

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Performance characteristics of the
South African Triage Scale
(Adult version)

By

Michèle Twomey

Thesis presented for the degree of
Doctor of Philosophy
in the School of Public Health and Family Medicine
University of Cape Town

August 2011

Performance characteristics of the
South African Triage Scale
(Adult version)

By

Michèle Twomey

Student No: TWMMIC001

BSc

Thesis presented for the degree of
Doctor of Philosophy
in the School of Public Health and Family Medicine
University of Cape Town

August 2011

Supervisors: Professors Jonny Myers, Lee Wallis & Mary Lou Thompson

This thesis is presented in fulfilment of the requirements for the degree of Doctor of Philosophy (PhD) in the School of Public Health and Family Medicine, Faculty of Health Sciences, University of Cape Town. The work on which this thesis is based is original research and has not, in whole or in part, been submitted for another degree at this or any other university. The contents of this thesis is entirely the work of the candidate, or in the case of multi-authored published papers, constitutes work for which the candidate was the lead author. The contribution of the candidate to included multi-authored papers is further delineated in the preface to the thesis and in the introduction to each included paper as appropriate.

Michèle Twomey

August 2011

Abstract

The South African Triage Scale (SATS) was first implemented as an emergency centre triage scale around the Western Cape Province in 2006. This thesis describes the performance characteristics of the SATS when used by different levels of health care workers in Emergency Centres around the Western Cape Province. While exploring the most appropriate methods for reliability and validity appraisal, this thesis demonstrates good performance characteristics, including reliability and validity of the SATS.

The results chapter of the thesis is presented in the form of published or submitted papers. The first paper explores the methodology of reliability studies and demonstrates that the SATS has acceptable inter- and intra-rater reliability by including a novel graphical method for illustrating reliability measures. The second paper demonstrates in real time that the SATS provides reliable ratings when used at a primary health care level. The third paper explores different methods of triage scale validation and the challenges of these with particular reference to a less developed country such as South Africa. The fourth paper demonstrates substantial criterion validity of the SATS when using a local panel of experts as a reference and demonstrates the practical application of SATS by including the novel use of over/under-prediction. The final paper builds consensus on a set of vignettes using a modified Delphi method and uses the vignettes as a reference standard to validate the construct of acuity as measured by the SATS. The reference vignettes are made freely available to the reader so that they may be used for further triage scale validation studies.

The thesis concludes that the SATS has good performance characteristics, which supports the feasibility of further implementation in similar settings. It is suggested that this emergency centre triage scale be endorsed and adopted at National Health care level.

Acknowledgements

I would like to express my sincere thanks and acknowledge my supervisors who have contributed enormously to this thesis:

- Jonny Myers, for his excellent guidance, insight and encouragement
- Lee Wallis, for his commitment, inspiration and vision to establish a triage system, and create the opportunity for me to become involved in 2005 in a project that would set the course of my work until now.
- Mary Lou Thompson, for her statistical expertise, insight and constant support

I would also like to express my deepest thanks to:

- The South African Triage Group and the Emergency Medicine Society of South Africa for supporting me
- All triage task teams, doctors and nurses at all facilities that have been actively involved in the implementation of the South African Triage Scale.
- My family, who always encouraged me, stood by me and believed in me.
- The loving support and patience of my partner, Gerry

Preface

This thesis includes published papers, as per general provision 6.7 in the General Rules for the Degree of Doctor of Philosophy (PhD) of the University of Cape Town, and with the approval in 2011 of the University Doctoral Degrees Board. The following five papers are formally included as part of the thesis:

1. Twomey M, Wallis LA, Thompson ML, Myers JE. The South African Triage Scale (adult version) provides reliable acuity ratings. *International Emergency Nursing*, in press.
2. Twomey M, de Sá A, Wallis LA, Myers JE. Inter-rater reliability of the South African Triage Scale: assessing two different cadres of health care workers in a real-time environment. *African Journal of Emergency Medicine*, in press.
3. Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. *Emergency Medicine Journal* 2007; 24: 477-479.
4. Twomey M, Wallis LA, Thompson ML, Myers JE. The South African Triage Scale (adult version) provides valid acuity ratings when used by doctors and enrolled nursing assistants. *African Journal of Emergency Medicine*, in press.
5. Twomey M, Wallis LA, Thompson ML, Myers JE. Validating the construct of acuity against a set of reference vignettes developed via modified Delphi method. *Emergency Medicine Journal*, in press.

The contribution of the candidate is discussed as an introduction to each paper (pages 17, 35, 45, 51 & 69). In summary, the candidate was the lead and corresponding author on all of the included papers, and drafted all versions of the manuscripts. All co-authors critically reviewed and approved the submitted manuscripts, and any comments were assessed by and where appropriate integrated by the candidate. The senior or a senior co-author on each paper has separately confirmed to the University of Cape Town Doctoral Degrees Board that the included papers overwhelmingly reflect the candidate's own scientific work. The candidate personally conducted all of the analyses in the included papers (as outlined in the methods sections of the papers).

An appendix details further papers in which the candidate is an author, and which have additionally been referenced in either the literature review or synopsis.

Abbreviations

ACSCOT	American College of Surgeons Committee on Trauma
ATS	Australasian Triage Scale
CTAS	Canadian Emergency Department Triage and Acuity Scale
CTG	Cape Triage Group
CTS	Cape Triage Score
EC	Emergency Centre
ED	Emergency Department
EN	Enrolled Nurse
ENA	Enrolled Nursing Assistant
ESI	Emergency Severity Index
ETAT	Emergency Triage Assessment and Treatment
FRENCH	French Emergency Nurses Classification in Hospital scale
HCW	Health Care Worker
ICC	Intraclass Correlation Coefficient
MEWS	Modified Early Warning Score
MTS	Manchester Triage System
NPV	Negative Predictive Value
PPV	Positive Predictive Value
QWK	Quadratically Weighted Kappa
SATG	South African Triage Group
SATS	South African Triage Scale
SHCP	Senior Health Care Professional
TEM	Triage Emergency Method
TEWS	Triage Early Warning Score
TTS	Taiwan Triage System

Table of Contents

CHAPTER 1: INTRODUCTION AND LITERATURE REVIEW	1
SECTION A: INTRODUCTION	1
<i>Emergency Centre triage</i>	2
SECTION B: BACKGROUND AND LITERATURE REVIEW	5
<i>Emergency Centre triage scales</i>	6
<i>Performance Characteristics</i>	8
Reliability	8
Terminology	8
Existing methods of measuring reliability	8
Limitations of the Kappa coefficient	9
Standards for the interpretation of the kappa coefficient	10
Measurement of reliability coefficients	10
Validity	11
Terminology & types of validation	11
Existing validation methods and measures for criterion and construct validation	13
Limitations of existing methods and measures of triage scale validation	13
<i>Development of SATS</i>	14
SECTION C: AIMS AND OBJECTIVES	15
SECTION D: METHODS AND ETHICS	16
<i>Methods</i>	16
<i>Ethical approval</i>	16
CHAPTER 2: RESULTS IN THE FORM OF PUBLISHED PAPERS	17
THE SOUTH AFRICAN TRIAGE SCALE (ADULT VERSION) PROVIDES RELIABLE ACUITY RATINGS	17
INTER-RATER RELIABILITY OF THE SOUTH AFRICAN TRIAGE SCALE: ASSESSING TWO DIFFERENT CADRES OF HEALTH CARE WORKERS IN A REAL-TIME ENVIRONMENT	35
LIMITATIONS IN VALIDATING EMERGENCY DEPARTMENT TRIAGE SCALES	45
THE SOUTH AFRICAN TRIAGE SCALE (ADULT VERSION) PROVIDES VALID ACUITY RATINGS WHEN USED BY DOCTORS AND ENROLLED NURSING ASSISTANTS	51
VALIDATING THE CONSTRUCT OF ACUITY AGAINST A SET OF REFERENCE VIGNETTES DEVELOPED VIA MODIFIED DELPHI METHOD	69
CHAPTER 3: DISCUSSION	89
RELIABILITY	89
<i>Limitations of the quadratically weighted kappa and how to overcome them</i>	89
<i>Limitations of the software packages</i>	90
VALIDATION	90
<i>Current methods of triage scale assessment and their limitations</i>	90
<i>How can limitations be overcome?</i>	91
<i>Delphi Methodology</i>	92
<i>Validation measures for triage scales</i>	92
<i>Consequences of mis-triage versus mis-prediction</i>	93
LIMITATIONS AND GAPS IN THE THESIS	95
<i>Reliability</i>	95
<i>Validation</i>	95
CONCLUSION AND RECOMMENDATIONS FOR FURTHER RESEARCH	96
APPENDIX: ADDITIONAL CO-AUTHORED PAPERS	97
REFERENCES	98

Table of Figures

Figure 1: Continuum of resource availability and corresponding variations in triage ⁷	2
Figure 2: Relationship between capacity and demand at triage.....	2
Figure 3: Illustration of over- and under-prediction of rater's assigned acuity categories as acuity distribution changes in different contexts A, B and C.....	94
Table 1: Time to treatment in four internationally developed 5-level triage scales.....	6
Table 2: Time to treatment in four internationally developed 4-level triage scales.....	7
Table 3: Various criteria for kappa	10
Table 4: Two hypothetical contexts and this study with different acuity distributions.....	94

University of Cape Town

Chapter 1: Introduction and literature review

Section A: Introduction

South Africa has a population of approximately 49 million people with an average life expectancy of around 51 years.¹ Human immunodeficiency virus, intentional and un-intentional injuries, communicable diseases such as tuberculosis and chronic diseases such as hypertension and diabetes mellitus all contribute to the unique quadruple burden of disease experienced by the 42 million South Africans that access public health care.² It is therefore not surprising that overcrowding is a frequent occurrence in public hospital emergency centres (ECs) in South Africa.³ Emergency centre visits are unscheduled and unpredictable in terms of arrival rates and patient severity. This has resulted in a need to determine guidelines for prioritising acutely ill patients and efficient use of resources: triage systems help EC staff to achieve this.

Triage originated in the eighteenth century and refers to the process of sorting for treatment priority in order to conserve manpower and increase the likelihood of survival.⁴ The chief surgeon of Napoleon's Imperial Guard, Baron Dominique Jean Larrey, is said to be the founder of the process of prioritising the wounded based on their need for surgical intervention rather than rank.⁵⁻⁶ Iserson and Maskop refer to three conditions that need to be met before the term "triage" can be used in the health care context: (i) resources are scarce relative to demand (ii) healthcare workers briefly assess patient's medical needs prior to intervention (iii) treatment priority is based on urgency using an established system.⁷

Over the last two centuries triage has been applied to a variety of healthcare settings and operates on a continuum.⁷ The far left extreme of the continuum as seen in **Figure 1** represents circumstances with (i) smallest number of available resources (ii) lowest resources to patient ratio and (iii) least social order. This signifies triage in a major incident (the locally accepted term for Disaster or Mass Casualty Incident) where the decision underlying the concept of urgency in triage aims to do the greatest good for the greatest amount of people.⁸ The far right extreme represents triage in the intensive care unit with (i) largest number of resources (ii) highest resources to patient ratio and (iii) most social order.⁷ Here individual patient care is the priority and the decision underlying the concept of urgency in triage aims to identify patients needing intervention based on clinical deterioration and severity. Different health measurement scales and early warning scores have been developed to suit the needs at different ends of the continuum.

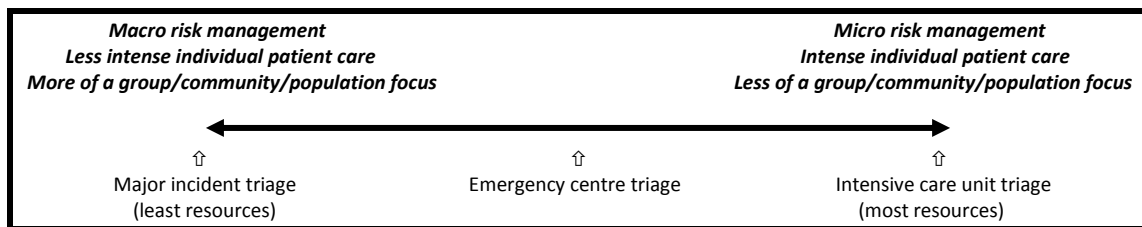


Figure 1: Continuum of resource availability and corresponding variations in triage⁷

Emergency Centre triage

EC triage is a system of clinical risk management that facilitates safer patient flow when clinical demand exceeds capacity.⁹ It will be found somewhere between the two extremes in **Figure 1** and is located more towards the right end of the continuum in developed countries and more towards the left end of the continuum in less developed countries. It is evident that triage becomes more of a public health issue as opposed to an individual patient care issue as circumstances change from right to left on the continuum.

Wherever one is across the globe (i.e. a remote rural field hospital in Afghanistan or an extremely advanced urban hospital in Australia), EC triage remains a dynamic process that shifts along the continuum of resource availability. It always functions simultaneously on a macro and micro risk management level with a heavier weighting on one or the other, dependant on resource availability (capacity) as well as number and acuity of patients presenting (demand). **Figure 2** illustrates that for each contextual variation in resource availability (i.e. number and level of skilled human resources, structure, equipment and stock), number and acuity of patients, EC triage will be unique and require an appropriate assessment of performance characteristics.

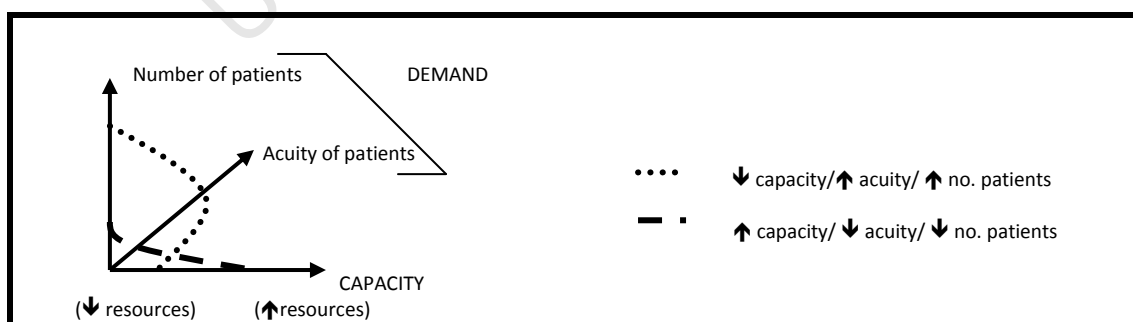


Figure 2: Relationship between capacity and demand at triage

Until recently South Africa had no triage system or formal process that sorted patients presenting to ECs according to urgency and treatment priority. In 2004 the Cape Triage Group (CTG), (now expanded and known as the South African Triage Group (SATG), developed a triage scale in an effort to standardize the triage process and improve patient care. The locally developed scale was

prospectively evaluated in the EC of an urban public hospital where it was shown to reduce mortality from 2% to 0.7%.¹⁰ Even though this study did not quantify the triage scales effect on morbidity, the results strongly suggest a patient's capacity to benefit from early recognition of urgency, further assessment and timeous treatment to prevent potential irreversible deterioration, discomfort and distress. In 2006 the Cape Triage Score (CTS), (now referred to as the South African Triage Scale (SATS)) was implemented as a policy for all ECs in the Western Cape Province.¹¹

It was during these initial years of rollout and implementation that this study was conceived, where the aim was to provide evidence of reliable and valid SATS use to support the feasibility of more widespread implementation. Over time the project has identified various methodological difficulties found in the process of evaluating triage scale reliability and validity, and study findings embrace the application of possible resolutions, including the use of standardized terminology and explicit methods in relation to the triage literature, that will guide future research.

University of Cape Town

Section B: Background and literature review

An extensive literature search was undertaken at the beginning of the research period (2007), and was repeated at regular intervals throughout completion of the thesis. The last search was undertaken in June 2011. A combination of electronic medical databases, electronic journals and medical library information was searched. Search terms varied for different sections of the thesis falling mainly into reliability of triage scales, validation of triage scales, methods in reliability and validity studies and Delphi methodology.

Electronic medical databases included:

- Medline (1970 – to date)
- PubMed Central
- Cinahl
- Proquest Dissertations and Theses
- Cochrane library (all databases)

Electronic journals included:

- Academic Emergency Medicine
- Accident and Emergency Nursing
- Annals of Emergency Medicine
- International Emergency Nursing
- Journal of Accident and Emergency Medicine
- South African Medical Journal
- Canadian Journal of Emergency Medicine
- Journal of Clinical Epidemiology
- Emergency Medicine Journal
- European Journal of Emergency Medicine

The Google® search engine and Google Scholar® were searched using more general terms. Hand searching of the following journals was undertaken at the University of Cape Town medical library:

- Accident and Emergency Nursing (2007 – 2011)
- British Medical Journal (1995 – 2011)
- South African Medical Journal (1995 – 2003)

Attempts were made to search the grey literature through sources including:

- The System for Information on Grey Literature in Europe
- Dissertation abstracts

The bibliographies of all articles retrieved were searched for any further relevant articles.

Emergency Centre triage scales

Globally, an increased demand for emergency healthcare combined with increasingly congested ECs and access block (whereby patients who are to be admitted cannot move to the wards due to underutilized ward beds, and therefore receive ongoing care in the EC) has led to an increased risk of mortality in the EC.¹²⁻¹⁶ This has resulted in a need for systems to ensure that limited available resources reach those in most urgent need. EC triage is a formal but dynamic process of rapidly and systematically determining a patient's acuity level so as to prioritize intervention/treatment and efficiently use EC resources.¹⁷⁻²¹ Different EC triage systems have been developed to suit a variety of contexts. In South Africa the four-level SATS has been widely adopted to facilitate the process of differentiating patient acuity levels.¹¹ The allocation of a triage category based on the SATS is primarily the function of a nurse. This requires a clinical decision that determines the patient's individual need for care. The process of triage ensures that emergency care is initiated in response to clinical need rather than order of arrival. It is essential that triage scales demonstrate the characteristics of good reliability and validity within the context of practice.²²⁻²³

Over the past two decades five 5-level triage scales have been developed, implemented and revised: the Australasian Triage Scale (ATS), (previously known as the National Triage Scale (NTS)),²⁴⁻²⁵ the Canadian Emergency Department Triage and Acuity Scale (CTAS),²⁶⁻²⁷ Emergency Triage (ET),⁹ the Emergency Severity Index (ESI),²⁸ and the French Emergency Nurses Classification in Hospital Scale (FRENCH).²⁹ The ATS, CTAS and FRENCH assign acuity based on the patient's presentation (i.e. each priority level includes various clinical descriptors and different vital signs or parameters).²⁵⁻²⁶ ET has 52 presentational flowcharts that guide the triage practitioner to the patient's clinical priority.⁹ The ATS, CTAS, ET and FRENCH have all been designed with a safe maximum 'time to treatment' for each triage level as summarised in **Table 1**. This has been revised in the latest version of the CTAS to the estimated 'time to reassessment' as initial time to treatment was not being met.³⁰

Acuity level	ATS	CTAS	ET	FRENCH
Level 1	Immediate	Immediate	Immediate	Immediate
Level 2	Within 10 min	Within 15 min	Within 10 min	Within 20 min
Level 3	Within 30 min	Within 30 min	Within 60 min	Within 60 min
Level 4	Within 60 min	Within 60 min	Within 120 min	Within 120 min
Level 5	Within 120 min	Within 120 min	Within 240 min	Within 240 min

Table 1: Time to treatment in four internationally developed 5-level triage scales

The ESI is different in that it has no 'time to treatment or re-assessment' associated with each triage level. It consists of one flowchart and besides rating the patient's acuity the triage nurse also estimates the number of resources (such as X-ray, laboratory tests etc.) a patient at level 3-5 is likely to require.¹⁷ There are several studies of how these triage scales have been adopted in other

countries and institutions including ET in the Netherlands,³¹⁻³² Sweden,³³ Hong Kong³⁴ and Ireland,³⁵ CTAS in Sweden³⁶ and Andorra,³⁷ ESI in Netherlands,^{32 38} Germany³⁹ and Greece⁴⁰ and ATS in Belgium.⁴¹ It is not clear whether they can be translated to less developed countries as they were produced and designed for developed countries.⁴²

Four 4-level triage scales have been developed, namely the South African Triage Scale (SATS),¹⁰ the Italian triage emergency method (TEM),⁴³ the Swiss triage scale⁴⁴ and the Taiwan Triage Acuity Scale (TTS).⁴⁵⁻⁴⁶ They all assess medical urgency based on physiologic parameters and a list of clinical discriminators. Similar to the ATS, CTAS, ET and FRENCH the 4-level scales incorporate target times to treatment. The Italian TEM has some similarities with the ESI in that the triage process requires the triage nurse to estimate the number of resources that the patient is likely to require.

Acuity level	SATS	TEM	TTS	Swiss triage scale
Level 1	Immediate	Immediate	Immediate	Immediate
Level 2	Within 10 min	Within 20 min	Within 10 min	Within 20 min
Level 3	Within 60 min	Within 60 min	Within 30 min	Within 120 min
Level 4	Within 240 min	Within 120 min	Delayed or OPD*	Delayed or OPD*

* Receives delayed treatment or may be seen in the outpatient department

Table 2: Time to treatment in four internationally developed 4-level triage scales

The rapid and widespread development of new EC triage scales calls for a comprehensive evaluation of their performance characteristics using proper statistical methods. There are undoubtedly several EC triage scales that are either currently in use and have not yet been evaluated and published or are in the process of being developed. There is no consensus in the emergency medicine arena whether to use a 3-, 4- or 5-level scale. While many countries and institutions still use a subjective 3- level scale (I=life-threatening, II=potentially life-threatening, III=non life-threatening), the emergency medicine triage literature has provided evidence for a shift towards more objective, systematic 4- and 5-level scales.^{18 42 47}

The key features common to all scales: (i) they are all ordinal rating scales (ORS), (ii) they require raters to assign a triage acuity level, (iii) they lack an objective criterion for use in validation of triage scoring. These features in turn also give rise to problems when assessing rating system performance characteristics such as reliability and validity.

Performance Characteristics

Reliability

It is important to first gather evidence of a scale measuring something in a reproducible fashion before evaluating whether it is measuring what is intended.⁴⁸ Reliability is not an inherent property of a scale; rather it is an interaction between the instrument, the group of people using the instrument, and the context.⁴⁸ It refers to the degree to which repeated assessments of the same patient with a triage scale will deliver the same acuity level. Inter-rater reliability determines the variability among multiple raters and intra-rater reliability assesses the variability within a single rater at different time intervals.⁴⁹ In the absence of an objective criterion, reliability is an important performance indicator for evaluating measurement quality and reducing error.⁴⁹ High inter-rater reliability is desired as it reduces variability in patient care where the instrument is the same and raters and contexts are similar.⁴⁹ Reliability determines how useful the ratings of a triage scale are, by reflecting the amount of random error inherent in any measurement.⁴⁸

Terminology

In the health care arena reliability is frequently referred to as 'precision', 'accuracy', 'repeatability', 'consistency', 'agreement', 'concordance' and 'reproducibility' among others.⁴⁸ While some authors may be describing a narrower aspect of reliability and correctly naming it 'repeatability', 'reproducibility' or 'consistency', the vast majority using these terms cannot agree amongst themselves what they mean, and often use them interchangeably.⁴⁸ This is problematic and causes confusion as some authors use 'accuracy' as a synonym for reliability; while others use it as a synonym for validity.⁴⁸ For the purpose of this study and for clarity the use of the term reliability is preferred, where the concept of reliability, does not refer to an inherent property of the rating scale or system, but rather to the inter- and intra-rater variability in the application of a particular scale by different raters in a particular context.⁴⁸

Existing methods of measuring reliability

Among numerous statistical approaches, no single measurement has been adopted as the standard for describing reliability.⁴⁹ Common measures used in reliability analyses include percentage agreement,⁵⁰ Cohen's kappa coefficient⁵¹ and the Intraclass Correlation Coefficient (ICC).⁵² The use of percentage agreement (i.e. exact agreement) on its own does not account for both percentage agreement and percentage agreement expected by chance.⁵³ The unweighted kappa coefficient corrects the observed percentage of agreements between the raters for the effect of chance.⁵³ There are however scales where disagreements between raters may not all be equally important. For

example, on an ordinal triage scale, there could be disproportionately greater implications if the two acuity categories chosen by the raters are further apart than one adjacent category. To account for the magnitude of disagreement, Cohen introduced the weighted kappa coefficient.⁵⁴ While many triage reliability studies incorporating weighted kappa do not specify the form of weighting, the linear weights introduced by Cicchetti and Allison⁵⁵ and quadratic weights introduced by Fleiss and Cohen are most commonly used.⁵⁶ Shrout and Fleiss define three types of ICCs;

- Type 1 assumes that a different set of raters is used for each vignette, or equivalently, that there is no correlation between ratings by the same rater across different vignettes.
- Type 2 assumes that the raters are a random sample from a population of raters
- Type 3 assumes a given number of raters and these are the only raters of interest for the reliability of these ratings.⁵⁷

Each type of ICC can be used to assess either:

- (i) the reliability of a single raters' ratings (among a group of k raters) or
- (ii) the reliability of the mean rating of the k raters (i.e. what is the reliability of a rating that is calculated by averaging the ratings across raters).⁵⁷

The quadratically weighted kappa estimates have been shown to be equivalent to the ICC (type 2,i).⁵³ In the case of two raters, the ICC (type 3,i) was shown to equal the Pearson's product-moment correlation coefficient between these two raters' ratings.⁵⁷ As a measure of reliability for ordinal scales however, the Pearson correlation has been regarded inappropriate and theoretically incorrect.⁴⁸

Both exploratory and inferential statistical procedures are needed to fully characterize reliability, and both can provide useful information to help improve reliability.⁴⁹ Nelson and Pepe suggest a novel exploratory method for ordinal data that describes measurements of variability graphically by examining frequency distributions of rating values.⁴⁹ This graphical method can provide a meaningful summary of inter-rater reliability that is easy to interpret and accommodates many raters.⁴⁹

Limitations of the Kappa coefficient

Kappa-type statistics have several limitations that have been discussed in the literature.^{49 58-61} Reliability expressed as a collapsed single summary statistic between -1 and 1, (with a negative value indicating poorer than chance agreement, zero indicating exact chance agreement and a positive value indicating better than chance agreement), may dilute important differences between raters found within each category of the ordinal scale.⁴⁹ The kappa statistic is mathematically dependant on the number of categories used in the triage scale (i.e. more categories give rise to higher values).^{49 58} Furthermore the distribution of cases may deflate the kappa statistic and lead to false conclusions of poor reliability (e.g. higher proportions of extreme cases will decrease variability because there is

more clinical agreement with extreme cases).⁴⁹ Thus interpretation of the kappa statistic is relative and not always as clear-cut as the recommended acceptable values which are discussed below.

Standards for the interpretation of the kappa coefficient

There are varying opinions on what magnitude of reliability is 'good enough'. While some acceptable values have been recommended, they are basically all arbitrary judgements.⁴⁸ Authors such as Landis and Koch, Fleiss, Cicchetti and Sparrow, have proposed various criteria for interpreting kappa values. These are summarised in **Table 3**.

Kappa	Landis & Koch ⁶³	Fleiss ⁵³	Cicchetti & Sparrow ⁶⁴
< or equal 0	Poor	Poor	Poor
0.0 – 0.20	Slight		
0.21 – 0.40	Fair		
0.41 – 0.60	Moderate	Fair to Good	Fair
0.61 – 0.75	Substantial		Excellent
0.75 – 0.80		Excellent	
0.80 – 1.00	Almost Perfect		

Table 3: Various criteria for kappa

Measurement of reliability coefficients

Exploratory and inferential statistical procedures are needed to fully characterize reliability. Both can provide important information to help improve reliability.⁴⁹ Following extensive debates in the literature regarding the most appropriate choice of reliability coefficient, all forms of the ICC appear to be the preference and superior to the weighted kappa when assessing reliability in ordinal scales with more than two raters. Berk outlined the following reasons:

...(a) the ability to isolate factors affecting reliability; (b) flexibility, in that it is very simple to analyze reliability studies with more than two observers and more than two response options; (c) we can include or exclude systematic differences between raters (bias) as part of the error term; (d) it can handle missing data (e) it is able to simulate the effect on reliability of increasing or decreasing the number of raters; (f) it is far easier to let the computer do the calculations; and (g) it provides a unifying framework that ties together different ways of measuring inter-rater agreement – all of the designs are just variations on the theme of the ICC.⁴⁸

This study reports and compares all of the above reliability summary statistics as well as the novel graphical method recommended by Nelson and Pepe.⁴⁹

Validity

Validity testing allows one to assess the acuity inferences made about patients being triaged and estimates how close to the truth those inferences are.⁴⁸ It refers to the extent to which a triage scale is actually measuring what it intends to measure independent of its reliability (i.e. is the triage scale correctly identifying the true urgency of patients).⁴⁸

The purpose of a scale does not necessarily determine how the levels of a scale are defined, what indicators are used to reach different levels on the scale, and how validity is assessed, as there are often many ways in which a scale is used.⁴⁸ Whether a scale can be used in a particular way depends on whether it was validated for that purpose.⁴⁸ While every triage scale has been developed with one universal purpose of prioritizing patients based on clinical urgency, a plethora of context specific underlying notions exist as to what constitutes urgency and different urgency levels. The process of validation implies that an external reference or gold standard measure of urgency is required. This constitutes a challenge considering that no single outcome measure captures the concept of urgency.⁶⁵ To overcome this, various reference standards have been selected for the validation process in past studies. Consensus is yet to be achieved on the most appropriate reference standard.

Terminology & types of validation

Accuracy and validity have in some instances been used synonymously, but may in fact be quite different depending on the definition of accuracy. If accuracy is defined as “distance from the truth” then referring to it as validity is acceptable, but if accuracy is defined as “precision” then referring to validity as accuracy is incomplete and inappropriate.⁶⁶ Many different types of validity have been introduced via textbooks, and defined in great detail.^{48 67} Streiner and Norman report on a reconceptualization of validity that reveals any validation to be a process of hypothesis testing.⁴⁸ This will prevent much confusion and wasted time deciding on the type of validity being assessed even though this conceptual shift may take several years to filter from the field of education and psychology to health care.⁴⁸ Following the suggestion by Streiner and Norman this study has chosen to refer to ‘validity’ as a unitary concept, but describes different types of validation processes.⁴⁸

The first type is content validation which refers to all aspects around triage scale construction including the physiological measures and clinical indicators that determine a particular level of urgency.⁴⁸ This type of validation is not based on the evaluation of ratings obtained from the scale, or performance differences between raters using the scale; it merely pertains to expert consensus regarding content.⁴⁸

The second type, criterion validation, refers to the correlation of a rating of urgency with the objectively determined truth regarding urgency as used and accepted in the field.⁴⁸ There is currently no used and widely accepted objectively determined truth for urgency in the field of emergency

medicine triage, and moreover there is also no widely agreed upon best reference standard.⁴⁸ Criterion validation can be divided into concurrent validation, where the scale is correlated with a criterion at the same time, and predictive validation where the scale is correlated with a criterion that will only be available in the future. It has been proposed that diagnostic utility and predictive utility are more descriptive and meaningful terms for concurrent and predictive validation respectively.⁶⁸

Construct validation is the third type of validation that pertains to abstract variables such as pain, anxiety and urgency. These cannot be directly observed, but hypothesized indicators and manifestations in terms of patients' behaviour may be.⁴⁸ The underlying indicators and manifestations are referred to as constructs that explain the relationship among behaviours and other variables.⁴⁸ When undertaking construct validation it is the construct of different urgency levels that one is testing which in turn is an indication of how close to the truth the inferences are that one makes about patients based on the ratings of a scale. It is important to note that when testing the construct of urgency, there are a number of further constructs underlying this very construct (e.g. pain is both a construct itself and a crucial component of another construct: urgency level). Understandably construct validation is an on-going process that differs methodologically from any other validation testing,⁴⁸ because there are various components, relationships and predictions that can be tested based on the theory of the construct, and there is no one single experiment that can explicitly verify a construct.⁴⁸ As an example one may choose to validate the construct of urgency as measured by a triage scale by developing a reference standard for urgency through an international consensus building method and then correlate this with ratings of urgency as used and accepted in the field. If a positive correlation is found between the local triage scale ratings and the international reference standard one can deduce that the construct of acuity as measured by that triage scale is valid.

Cronbach and Meehl outlined three mandatory steps for construct validation:

- (i) unambiguously spell out a set of theoretical concepts and how they are related to each other
- (ii) develop scales to measure these hypothetical constructs
- (iii) actually test the relationships among the constructs and their observable manifestations.⁶⁹

Clark and Watson recommend a brief, written description of the construct, how it will manifest itself objectively, how it is related to other constructs and behaviours and what theory will be tested before embarking on any construct validation. Furthermore they state the following:

This emphasis on theory is not meant to be intimidating. That is, we do not mean to imply that one must have a fully articulated set of interrelated theoretical concepts before embarking on scale development. Our point, rather, is that thinking about these theoretical issues prior to the actual process of scale construction increases the likelihood that the resulting scale will make a substantial contribution to the literature.⁶⁹

When estimating validity, it should not be the validity of a scale that is reported on, instead it should be the ratings obtained from a scale and the inferences made about a sample population when a scale is used for a particular purpose.⁴⁸

Existing validation methods and measures for criterion and construct validation

Measures of association between urgency levels and various outcome markers such as admission, death, resource utilisation, length of stay and diagnosis are frequently reported in the triage literature.^{10 17 29 37 40 69-74} Validation processes also commonly report the kappa statistic and sensitivity analyses, where percentage sensitivity indicates the probability of triaging the patient correctly using expert consensus as an objectively determined truth regarding urgency.^{10 29 31-33 36 43-44 47 72 76-77} Associated over- and under-triage refers to the degree of inevitable mis-triage found in any triage scale. Under-triage occurs when the patient's acuity level is recognised as less urgent than it truly is, which may lead to longer waiting times, delayed treatment and an increased risk of patient mortality and morbidity. Over-triage occurs when the patient's acuity level is recognised as more urgent than it truly is, which in a resource poor setting where need exceeds capacity, may lead to inappropriate use of resources and again increased risk of patient mortality and morbidity.⁷⁸

Moll states that triage studies may benefit from the methodology developed in diagnostic research.⁶⁵ Sensitivity, specificity and percentage over-/under-triage are accuracy summaries (accuracy defined here as the distance from the truth). These summaries may be compared across studies and do not depend on the acuity distribution in a given setting. By contrast, predictive values provide information on actual performance in a given setting. In practice, when a patient is assigned a particular acuity level, the true acuity is not known, and what is of interest from the patient care and resource management perspective is how likely that assigned acuity is to be correct. Predictive values indicate the proportion of patients with high acuity levels who are correctly identified as high acuity. The purpose of having a triage scale is to use it to prioritize high acuity patient who need care the soonest, so it is crucial to establish the probability that the scale will correctly identify the highest acuity patients. Sensitivity and specificity will not give us this information, but approaching the data from the direction of scale ratings, using predictive values will.⁷¹ One study has included predictive values as additional measures of performance characteristics.⁴³ The inherent limitation regarding predictive values is that they are dependent on prevalence, and when considering them for triage scales the patient sample population should reflect the prevalence of acuity levels in the population which may be extremely variable and difficult to achieve.

Limitations of existing methods and measures of triage scale validation

There is variability across different countries and EC contexts regarding the development of triage scales, their application, the construct of urgency, the triage validation process and which outcome marker/s to use as a proxy gold standard. It is not clear what validation measures are the most

appropriate and to what degree a measure reflects valid results. Reference guidelines on the acceptable ranges of mis-triage in EC triage could not be found. The American College of Surgeons Committee on Trauma (ACSCOT), have however provided acceptable triage rates for pre-hospital triage of trauma patients. An average rate of 30-50% over-triage is considered acceptable (and in some cases necessary) to prevent under-triage exceeding 10-15%.⁷⁹ These ranges were developed for pre-hospital triage of trauma cases and not for in-hospital trauma and medical cases. A recent retrospective cohort study indicated that achieving these ACSCOT benchmark was not feasible in Pennsylvania, and that these guidelines needed modification if they were to be implemented.⁸⁰ In spite of this the ACSCOT guidelines are sometimes used as a reference for interpreting average over- and under-triage for EC triage in the emergency medicine literature.

Development of SATS

The SATG, consisting of emergency physicians, nurses and paramedics, obtained permission to borrow the Modified Early Warning Score (MEWS) and contextually adapt it to South Africa. The physiological parameters of the MEWS were modified to reflect the ranges appropriate to South Africans, and the algorithm was renamed the Triage Early Warning Score (TEWS)⁸¹. A list of clinical discriminators was established, and via expert consensus, the content of the SATS was finalised in such a manner that specific high-risk indicators and criteria were aligned to respective acuity levels and target times to treatment for practical application.⁸¹

Prospective evaluations were undertaken at an urban public and private hospital in 2005 after which the triage instrument was developed, tested, amended and again tested more widely. Outcome markers chosen for the two studies were admission, discharge and mortality. Measures of association were calculated to test the hypotheses that the higher acuity levels predicted a higher probability of admission and that the low acuity category predicted a higher probability of discharge. Both studies reported this to be the case and following specific amendments in the public hospital study reduced the under-triage rate from 24% to 12%.¹⁰

Moll highlights that in using the methodology of diagnostic research, there are steps to validating a triage system of which the first is content validation (i.e. consensus based derivation of the triage scale for the different urgency levels).⁶⁵ The second step refers to criterion validation, which is an internal validation of the scale with a reference standard that is a best proxy in that setting.⁶⁵ Modifying the triage scale to improve urgency classification is the third step, and lastly the fourth step (construct validation) is the external validation in various emergency settings.⁶⁵ The SATS has been through steps one to three and the literature has indicated a need for validation in additional healthcare settings as well as an appraisal of the reliability of triage ratings obtained when using the scale.^{10 82-83}

Section C: Aims and Objectives

The above review outlines the concepts of reliability and validity and describes the process by which the SATS was developed and previously evaluated. It also highlights some of the methodological challenges one is confronted with in reliability studies and the validation process.

AIM: The aim of this thesis at the time of conceptualisation was to assess the performance characteristics of the SATS in detail.

PURPOSE: The purpose is to provide an informative summary of evidence supporting national adoption and implementation.

Specific objectives were:

1. to estimate the inter and intra-rater reliability of triage acuity ratings on EC patients by South African nurses and doctors using the SATS at secondary and tertiary level care.
2. to estimate the inter-rater reliability of individual versus team triage ratings by South African nurses and and final year medical students using SATS in a real-time environment.
3. to describe the current challenges in triage scale validation with a particular focus on less developed countries.
4. to validate SATS ratings by South African nurses and doctors against a reference standard.
5. to expand the scale's performance characteristics by introducing the terms over- and under-prediction which add a new dimension to mis-triage, pertaining to the scale's application in a given setting.
6. to validate the construct of acuity as measured by the SATS against an international expert panel as a reference standard.
7. to develop a set of vignettes that may be used as a reference standard for other validation studies.

Section D: Methods and Ethics

Methods

A series of vignettes were used for the reliability study and validation process. These were collected prospectively rather than retrospectively due to poor record keeping within facilities, and the limitations this would have on the extraction of information from patient folders. The 100 adult vignettes were abstracted from randomly selected actual EC case presentations at a busy urban hospital seeing 40,000 cases per year in the EC. Vignettes included gender, age, presenting complaint, mode of arrival, respiratory rate, heart rate, systolic blood pressure, level of consciousness and presence of injury.

A cross-sectional reliability study was undertaken using the series of 100 vignettes. Five different quantitative inter-rater reliability measures with 95% confidence intervals as well as exploratory graphical displays were presented.

In addition a cross-sectional reliability study was undertaken in a real-time environment with 50 patients. Inter-rater reliability measures with 95% confidence intervals were presented for individual and collaborative team triage.

The cross-sectional validation study used the same series of 100 adult vignettes, and an objective reference standard was established by obtaining consensus from 2 experts on the acuity ratings for each of the 100 vignettes. Ratings obtained by doctors and nurses using the SATS were validated against the proxy reference standard ratings. Measures including sensitivity, specificity, under- and over-triage, predictive values with over- and under-prediction were reported.

A 2-round modified Delphi method was conducted to obtain 80% consensus from an international panel of 18 triage experts on the previously mentioned series of 100 adult vignettes.

Ethical approval

Ethical approval was obtained from the University of Cape Town Research Ethics Committee (REC REF:063/2005). This approval covered all analyses included in this thesis. The approval did not require individual informed consent, as it was based on the premise that all clinical care was standard of care, that the research component consisted solely of ensuring completeness of routine data, and that data would be kept anonymous at all times. All three of these requirements have been continuously met.

Chapter 2: Results in the form of published papers

The South African Triage Scale (Adult Version) provides reliable acuity ratings

Paper overview

This paper presents the inter- and intra-rater reliability of the South African Triage Scale when used by both emergency physicians and enrolled nursing assistants (ENAs) on a general adult EC patient population in South Africa. It also considers the most appropriate method for assessing reliability for EC triage scales.

Contribution to the thesis and novelty

This paper addresses the first objective and one of the important components of any triage scale. There is currently no other study that assesses the reliability of the SATS. Other contributions of the paper that are unique are the use of South African vignettes and the use of a graphical method to illustrate inter-rater variability and provide a more informative reliability summary.

Contribution of candidate

The candidate was responsible for data management and conducted all the analysis in the manuscript. The candidate wrote and managed all drafts of the manuscript, incorporating suggestions from co-authors.

Publication status

The paper has been accepted 28 July 2011 for publication in International Emergency Nursing.

Title:

The South African Triage Scale (Adult Version) provides reliable acuity ratings

Authors:

M.Twomey¹, LA. Wallis², M. Thompson³, J. Myers¹

Institutions:

- 1) School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
- 2) Division of Emergency Medicine, Department of Surgery, University of Cape Town & Stellenbosch University, Cape Town, South Africa
- 3) Department of Biostatistics, University of Washington, Seattle, United States

Word Count

Abstract: 193

Paper: 3329

University of Cape Town

Abstract

Objective: To estimate the inter- and intra-rater reliability of triage ratings on emergency centre patients by South African nurses and doctors using the South African Triage Scale (SATS).

Methods: A cross-sectional reliability study was performed. Five emergency physicians and ten enrolled nursing assistants independently assigned triage categories to 100 written vignettes unaware of the ratings given by others. Four different quantitative reliability measures were calculated and compared. Graphical displays portrayed rating distributions for vignettes with mean ratings at different acuity categories.

Results: The estimated quadratically weighted kappa for the group of emergency physicians was 0.76 (95% CI: 0.67-0.84) and for the group of nurses 0.66 (95% CI: 0.58 – 0.74). These values were close to the estimated intra-class correlation coefficients. For intra-rater reliability, the average exact agreement was 84%. The graphical displays showed that the least variability was evident in the vignettes that had a mean rating of 'emergency', 'very urgent' or 'routine'.

Conclusion: This study indicates good inter- and intra-rater reliability among nurses and doctors using the SATS. It suggests that the SATS is reliably applied, and supports the feasibility of further implementation of the SATS in similar settings.

INTRODUCTION

The South African Triage Scale (SATS), (previously named the Cape Triage Score (CTS)) is an initial measure of patient acuity in the Emergency Centre (EC) that was developed by the South African Triage Group (SATG) (Wallis, 2006), (previously named the Cape Triage Group (CTG)) (**Appendix 1**). Similar to the Manchester Triage System (MTS) (Mackway-Jones, 1997) and the Canadian Emergency Department Triage and Acuity Scale (CTAS) (Beveridge, 1998), the SATS incorporates target times to treatment. Patients are categorised into one of four acuity levels: Red (emergency – should be seen immediately), Orange (very urgent – should be seen in less than 10 minutes), Yellow (urgent – should be seen in less than 60 minutes), Green (routine – should be seen in less than four hours) and Blue (dead on arrival – should be certified within 2 hours)(Gottschalk et al., 2006). Resuscitation would always be initiated for a Blue patient, and an emergency physician would certify the patient dead. The SATS was intentionally designed for use by an Enrolled Nursing Assistant (ENA) due to the limited numbers of doctors and professional nurses in South Africa (Wildschut and Mqolozana, 2008). ENAs are entry-level nurses that have qualified with a one-year certificate (Subedar, 2005), and previous studies have provided some evidence of ENA competence using the SATS (Bruijns et al., 2008). The SATS has been implemented, monitored and refined in public ECs over the past 24 months (Bruijns et al., 2008); however, to date no study has assessed the reliability of SATS in South African Emergency Centres.

BACKGROUND

The reliability of a triage scale is a measure that tells us how standardised the application of a triage scale is. It refers to the agreement between different raters triaging a patient, and within a particular rater, triaging the same patient (Streiner and Norman, 2008). A reliable triage scale will result in different raters obtaining a similar acuity rating for a given patient. Triage tools should be highly reliable. Validity, on the other hand, assesses whether the triage scale correctly identifies the *true* acuity of the patient (Streiner and Norman, 2008) and is a characteristic distinct from reliability. In order to assess validity, an objective reference standard for the best measure possible of acuity is needed e.g. from a panel of experts. Before assessing the validity of a triage scale it is essential to provide evidence of its reliability (Worster et al., 2004). Much has been published in the emergency medicine literature on reliability and validity, however, these two concepts have not always been kept conceptually distinct. Confusion may arise where definitions of reliability and validity differ from the typical usage in epidemiology (van der Wulp and van Stel, 2009). Moreover, it is difficult to draw comparisons between different triage scales, as there is neither consensus nor clarity in the triage literature as to the most appropriate measure for assessing triage reliability. In addition scales have a different number of acuity categories and 4-level scales cannot be compared to 5-level or 3-level scales.

A literature review revealed 26 studies that estimated the reliability (defined as inter- and intra-rater agreement on triage acuity ratings) of 4- and 5-level triage scales by measuring the kappa coefficient and focusing on an adult population. Kappa coefficients are frequently used to quantify the level of agreement between multiple ratings of urgency categories in ordinal triage scales (Cohen, 1968). With ordinal scales, weighted rather than unweighted kappa coefficients are commonly used to weight disagreements according to magnitude of discrepancy (Cohen, 1968). Linear weights are proportional to the deviation of individual ratings and quadratic weights are proportional to the square of the deviation of individual ratings (Brenner and Kliebsch, 1996). Of the 26 studies reviewed, the weighted kappa was the most frequently reported statistic; however 19 (more than two thirds) of these studies did not report the specific type of weight used (e.g. whether linear or quadratic). Quadratic weighting was most frequent among studies that specified the weights used. These included only studies of 5-level triage scales. Other less frequently reported measures were the linear weighted kappa, percentage agreement between raters and the intra-class correlation coefficient (van der Wulp and van Stel, 2009).

The most common method of assessing reliability appears to be a weighted kappa statistic because it takes into account those disagreements within ordinal data that may warrant some (as opposed to zero) credit. The quadratically weighted kappa has been shown to approximate the intra-class correlation coefficient (Fleiss and Cohen, 1973). Kappa statistics however, whether weighted or unweighted, have a number of limitations when estimating reliability in that they are:

- dependent on the number of categories in the ordinal data (an increase in number of categories, increases the weighted Kappa estimate) (Brenner and Kliebsch, 1996).
- dependent on the distribution of cases (higher proportions of extreme cases will decrease the variability because there is typically more clinical agreement with extreme cases, and disagreement can only occur in one direction, towards the centre of the distribution)(Nelson and Pepe, 2000).
- insensitive to the differences in agreement for the different ordinal values (kappa values can hide important differences in agreement by creating a one-dimensional overall view that limits generalisability) (Nelson and Pepe, 2000).

The aims of this study are to estimate the inter- and intra-rater reliability of the SATS when used by both emergency physicians and ENAs on a general adult EC patient population in South Africa, and to consider the most appropriate method for assessing reliability.

METHODS

Study design & sample

A cross-sectional reliability study was conducted using a series of vignettes that had been collected prospectively. Prospective collection was favoured above retrospective collection due to the reality

that poor record keeping in patient folders would have limited the information to be extracted for vignettes. Adult vignettes were used as a suitable proxy for live triage cases (Worster et al., 2007). Based on their use in other studies (Considine et al., 2004, Olofsson et al., 2009, Worster et al., 2004) and the advantage of saving cost and time (Worster et al., 2004, Worster and Fernandes, 2005), this method is particularly useful in a less developed country such as South Africa. 100 adult vignettes were abstracted from randomly selected actual EC case presentations at Victoria Hospital, Cape Town (a busy urban hospital seeing 40,000 cases per year in the EC). From these 100 vignettes, 10 were randomly selected, duplicated and interspersed among the 100 to determine intra-rater reliability. The total of 110 vignettes included gender, age, presenting complaint, mode of arrival and vital signs. **Appendix 6** shows examples. Some vignettes also include information from further investigations such as blood glucose test and haemoglobin, which are done at triage. All information that was found in the triage paperwork was included in the vignettes.

Implementation of the SATS in the Western Cape province of South Africa began with a one-day mandatory training workshop for fifty ENAs and 25 emergency physicians. Five emergency physicians (20%) and 10 ENAs (20%) who completed their training were selected to represent different geographically located health facilities in the province. Ethical approval was obtained by the Research Ethics Committee, University of Cape Town (REC REF: 063/2005).

Triage scoring

Training for SATS implementation included a simple five-step triage approach. The initial two steps take a brief history and document the values of vital signs when a patient presents to the EC. Step 3 calculates a Triage Early Warning Score (TEWS) based on an age appropriate physiological scoring system (top section of **Appendix 1**). The patient is categorized in step 4 into an initial colour code based on the TEWS. For the purposes of this study the blue category – dead on arrival was not included. In step 5 the discriminator list (a safety net) is observed (bottom section of **Appendix 1**), allowing the patient to be upgraded into a higher priority colour code depending on whether their presenting complaint matches a discriminator in a higher priority colour category.

SATS Category	Target time to be seen by senior health care professional	Description of category	Coding for data entry
RED	Immediate	Emergency	0
ORANGE	Within 10 min	Very Urgent	1
YELLOW	Within 60 min	Urgent	2
GREEN	Within 240 min	Routine	3

Table 1: South African Triage Scale Categories (5)

This study assessed the reliability of the SATS in a general EC adult population. ENAs used the SATS to independently triage written sets of vignettes into one of 4 triage categories (**Table 1**), excluding the category dead on arrival. This was done in a classroom situation while emergency physicians accessed the vignettes online. In total 15 raters assessed 100 vignettes (1500 assessments) and 10 additional random duplicate vignettes for intra-rater reliability (150 assessments).

Data analysis

Reliability in this study was assessed by calculating several reliability measures to compare results of different measures within the study. The reliability measures considered in this study were unweighted, linear weighted and quadratically weighted kappa statistics, the intra-class correlation coefficient (Jakobsson and Westergren, 2005) and percentage intra-rater agreement. Inter-rater reliability with 95% confidence intervals was assessed for each professional group separately. Kappa coefficients were interpreted using the classification model designed by Landis and Koch (Landis and Koch, 1977). Graphical displays were also constructed, showing rating distributions for vignettes with a mean rating of a particular acuity category (Nelson and Pepe, 2000), hence having the ability to show differential agreement at different acuity categories, rather than summarising agreement in a single number, as is the case with the kappa statistic. Intra-rater reliability was assessed for each individual rater using percentage agreement.

These analyses were performed using AGREE 7 for Windows (version 7.002, 1999) to calculate the unweighted, linearly weighted and quadratically weighted kappa statistics, STATA statistical software package version 9.2, (version 9.2, 2006) to calculate the unweighted kappa statistic (with normal confidence intervals) and the intra-class correlation coefficient and to illustrate reliability via graphical displays and SPSS (version 17.0.0, 2008) to calculate the intra-class correlation coefficient.

RESULTS

Quadratically weighted kappa estimates were 0.66 (95% CI, 0.58 – 0.74) for ENAs and 0.76 (95% CI, 0.67 – 0.84) for emergency physicians. These values were in good agreement with the estimated intra-class correlation coefficient, which was 0.66 (95% CI, 0.60 – 0.73) for ENAs and 0.76 (95% CI, 0.70 – 0.82) for emergency physicians. Within individual raters of the ENA and emergency physician group the intra-rater reliability was substantial to almost perfect. Average exact percentage agreement for ENAs was 86% and for emergency physician 80%. The average percent agreement estimated to include a 1-level disagreement was 99% among ENAs and 96% among emergency physicians. **Table 2** summarizes all reliability measures calculated for this study.

Inter-rater reliability Measure	Profession Group	Point estimate (95% CI)
Unweighted kappa statistic	Nurses	0.51 (0.44 – 0.58)
	Doctors	0.68 (0.61 – 0.76)
Linearly weighted kappa statistic	Nurses	0.58 (0.51 – 0.65)
	Doctors	0.72 (0.64 – 0.79)
Quadratically weighted kappa statistic	Nurses	0.66 (0.58 – 0.74)
	Doctors	0.76 (0.67 – 0.84)
Intra-class Correlation coefficient	Nurses	0.66 (0.60 – 0.73)
	Doctors	0.76 (0.70 – 0.82)
Intra-rater reliability Measure	Profession Group	Mean (Min/Max)
Exact percentage agreement	Nurses	86% (70% - 100%)
	Doctors	80% (50% - 100%)
Percentage Agreement including a 1 category discrepancy	Nurses	99% (90% - 100%)
	Doctors	96% (80% - 100%)

Table 2: Summary of triage inter- and intra-rater reliability measures

Figure 1 (a-d) illustrates the distribution of combined nurses and doctors ratings for vignettes with a mean rating of emergency, very urgent, urgent and routine acuity respectively. The least variability in distribution of ratings can be seen in the vignettes that have a mean rating of emergency and routine acuity, i.e. at the extremes of the scale. **Figure 1 (a)** shows the rating distribution for 6 vignettes with a mean acuity category 'emergency', 90% of ratings were 0 (emergency) and the remainder of the ratings were only one category below that.

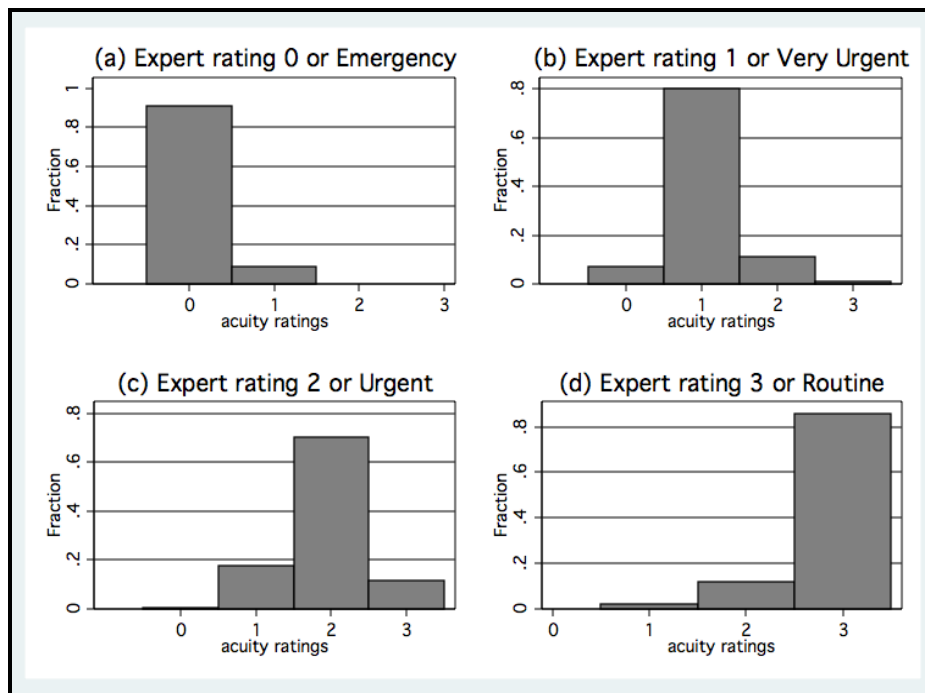


Figure 1 (a-d): Distribution of ratings for vignettes with a mean rating of (a) emergency (b) very urgent (c) urgent (d) routine

Figure 1 (c) shows the combined doctors and nurses rating distribution for 35 vignettes with a mean acuity category 'urgent', 70% of ratings were 2 (urgent), 18% of the ratings were one category above that (i.e. 'very urgent') and the remainder of the ratings were only one category below that (i.e. 'routine').

DISCUSSION

Acceptable inter- and intra-rater reliability was observed for emergency physicians and ENAs indicating that the SATS provides reliable triage acuity ratings among both groups. This demonstrates that the SATS is understood after a one-day training workshop and reliably applied by emergency physicians and ENAs alike.

Linear and quadratically weighted kappa measures for multiple raters cannot currently be calculated using STATA or SPSS. Based on reports and recommendations in the literature, AGREE 7 for Windows appeared to be an alternative that allowed us to do so (van der Wulp and van Stel, 2009). We found the package difficult to use and this was aggravated by the limited availability of adequate documentation and user support.

The estimated inter-rater agreement (measured as quadratically weighted kappa) among ENAs (0.66) and emergency physicians (0.76) was substantial. The average intra-rater agreement within each individual ENA (86%) and emergency physician (80%) was almost perfect.

Figures 1 (a-d) confirm the phenomenon that a higher frequency of extreme cases provides increased clinical agreement, highlighting the limitations of the kappa statistic (Nelson and Pepe, 2000) which provides a general estimate of agreement over all categories of severity, losing information about agreement for each specific category, and consequently limiting generalisability across study settings (Brenner and Kliebsch, 1996, Nelson and Pepe, 2000). The graphical approach allows an assessment of agreement for each specific category and hence facilitates generalisability across study settings.

Recently a novel adjusted weighted kappa for severity of mis-triage was proposed as an alternative, more appropriate, measure for triage reliability studies (van der Wulp and van Stel, 2009). The weighting scheme in this measure was designed to take into consideration the severity of over-triage and under-triage so as to reflect clinical triage practice. The authors introduce differential weightings for over- vs. under-triage based on the quadratically-weighted kappa (van der Wulp and van Stel, 2009). Over- and under-triage can, however, only be defined in relation to an objective reference and hence this adjusted weighted kappa was not considered to be a reliability measure. The adjusted weighted kappa for severity of mis-triage is a measure of *validity* that uses different standards, thereby limiting its utility for comparisons between validation studies. It cannot be used for assessing reliability, which is the focus of our study.

Reference	Triage Scale	Raters	Quadratically weighted Kappa (95% CI)	Linear weighted Kappa (95% CI)	Unweighted Kappa (95% CI)	Intra-class correlation coefficient (95% CI)
Beveridge et al., 1999 (17)	CTAS	Doctors & Nurses	0.80 (0.79-0.81)	-	-	-
		Doctors	0.83 (0.81-0.85)	-	-	-
		Nurses	0.84 (0.83-0.85)	-	-	-
Worster et al., 2004 (18)	CTAS	Nurses	0.91 (0.90-0.99)	-	-	-
	ESI	Nurses	0.89 (0.88-0.99)	-	-	-
Manos et al., 2002 (19)	CTAS	Nurses	0.80 (0.76-0.84)	-	-	-
		Physicians	0.82 (0.78-0.86)	-	-	-
		Basic Life Support (BLS)	0.76 (0.72-0.80)	-	-	-
		Advanced Life Support (ALS)	0.73 (0.68-0.77)	-	-	-
		Nurses, physicians, BLS & ALS	0.77 (0.76-0.78)	-	-	-
van der Wulp et al., 2008 (20)	MTS	Nurses	0.62 (0.60-0.65)	-	0.48 (0.45-0.50)	0.75 (0.72-0.77)*
Storm Versloot et al., 2009 (21)	MTS	Nurses	0.82 (0.74-0.89)	-	0.76 (0.68-0.83)	-
	ESI	Nurses	0.73 (0.64-0.83)	-	0.46 (0.37-0.55)	-
Olafsson et al., 2009 (22)	MTS	Nurses	0.81	0.71	0.61 (0.57-0.65)	-
Dong et al., 2006 (23)	eTRIAGE	Nurses	0.66 (0.60-0.71)	0.52 (0.46-0.57)	-	-
This study	SATS	Nurses	0.66 (0.58-0.74)	0.58 (0.51-0.71)	0.53 (0.51-0.56)	0.66 (0.60-0.73)
		Emergency Physicians	0.76 (0.67-0.84)	0.72 (0.71-0.92)	0.68 (0.62-0.74)	0.76 (0.70-0.82)

Table 3: A summary of seven reliability studies and their reported reliability measures

(*This point estimate refers to intra-rater agreement)

While the quadratically weighted kappa statistic is the most commonly reported statistic in the triage assessment literature, the literature reports that the kappa statistic suffers from the disadvantages of its dependence on the number of categories in the ordinal data, the frequency distribution of the mix of cases by severity, and it being a general estimate of agreement over all categories of severity losing information about agreement for each specific category. This limits generalisability across study settings (Brenner and Kliebsch, 1996, Nelson and Pepe, 2000).

Summary measures of reliability, particularly the kappa statistic, inadequately adjust for differential study characteristics, making meaningful comparison across reliability studies problematic. The above literature review highlighted seven of the 5-level triage reliability studies that elaborated on the type of weighted kappa used. The results of this current study, using a 4-level scale, cannot readily be compared to the results of 5-level triage scale but the range of values in **Table 3** provide some sense of reliability for each study in relation to the Landis and Koch benchmarks. We found that the SATS measures were similar and in one instance markedly better than the benchmarks.

LIMITATIONS

A limitation of any study assessing the reliability of triage scales using vignettes is that many nonverbal and contextual cues may be missed despite the fact that written vignettes are a very cost efficient way of assessing reliability (Worster and Fernandes, 2005, Considine et al., 2004). We tried to overcome the lack of qualitative nonverbal cues by leaving the subjective information in the description of vignettes, while keeping the patient information anonymous.

Intra-rater reliability was assessed within the same set of vignettes by duplicating 10 vignettes to create ten additional patient scenarios. There was no time interval before raters triaged the 10 additional vignettes, thus raters may have identified similarities. However the repeated vignettes were randomly distributed into the sets, which had to be completed in a certain order, and raters could not go back to previous triage ratings to compare differences.

There is a general limitation to all studies that attempt to compare reliability studies with the kappa because the comparison that is being made is not adequately standardised. We tried to overcome this by reporting the intra-class correlation coefficient as well as presenting graphical displays of the variability at each acuity level.

CONCLUSION

This is the first reliability study of a South African triage instrument using South African vignettes and an exploratory graphical approach. The results of this study have very important practical implications for the use of SATS in less developed countries such as South Africa. The reliability of SATS among ENAs and emergency physicians as a mixed group and as independent groups indicates favourable results especially when using the exploratory graphical approach to assess in more detail the variability within the four different urgency levels.

With a large amount of triage literature in the public arena there is still neither consensus nor clarity as to the most appropriate measure for assessing triage reliability; nor can there be given the complexity of the exercise in its varying contexts. In addition there is a paucity of published emergency medicine triage literature on the conceptual distinction between reliability and validity.

Graphical methods have been applied for the first time to examine inter-rater reliability for Emergency Centre triage tools. This provides useful additional information relating to specific triage urgency categories over and above the single measure of agreement over all urgency categories provided as a summary measure by a kappa statistic. Triage tool reliability assessment may benefit from this type of visual presentation, especially when comparing agreement around specific urgency categories between study settings. This study illustrates acceptable reliability of the SATS whether it is used by ENAs or emergency physicians, and provides supportive evidence for further implementation in similar settings by health professionals with different levels of training.

Publications on triage scale reliability should (i) clearly focus on reliability as distinct from validity (ii) have a clear description of the methods used (iii) include the graphical approach to provide a more informative reliability summary.

FUNDING- None

CONFLICT OF INTEREST - The authors declare that there is no conflict of interest

REFERENCES

- Beveridge, R., 1998. The Canadian triage and acuity scale: a new and critical element in health care reform. CAEP issues. Canadian Association of emergency physicians. *Journal of Emergency Medicine* 16, 507-511.
- Brenner, H., Kliebsch, U., 1996. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 7, 199-202.
- Bruijns, S.R., Wallis, L.A., Burch, V.C., 2008a. A prospective evaluation of the Cape triage score in the emergency department of an urban public hospital in South Africa. *Emergency Medicine Journal* 25, 398-402.
- Bruijns, S.R., Wallis, L.A., Burch, V.C., 2008b. Effect of introduction of nurse triage on waiting times in a South African emergency department. *Emergency Medicine Journal* 25, 395-397.
- Cohen, J., 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 70, 213-220.
- Considine, J., LeVasseur, S.A., Villanueva, E., 2004. The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Annals of Emergency Medicine* 44, 516-523.
- Fleiss, J.L., Cohen, J., 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as a measure of reliability. *Educational and Psychological Measurement* 33, 613-619.
- Gottschalk, S., Wood, D., DeVries, S., Wallis, L.A., Bruijns, S.R., 2006. The Cape Triage Score: a new triage system for South Africa. Proposal from the Cape Triage Group. *Emergency Medicine Journal* 23, 149-153.
- Jakobsson, U., Westergren, A., 2005. Statistical methods for assessing agreement for ordinal data. *Scandinavian Journal of Caring Science* 19, 427-431.
- Landis, J.R., Koch, G.G., 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 159-174.
- Mackway-Jones, K., 1997. *Emergency Triage*. Manchester Triage Group. BMJ Publishing Group.
- Nelson, J., Pepe, M., 2000. Statistical description of interrater variability in ordinal ratings. *Statistical Methods in Medical Research* 9, 475-496.

Olofsson, P., Gellerstedt, M., Carlstroem, E., 2009. Manchester triage in Sweden – interrater reliability and accuracy. *International Emergency Nursing* 17, 143-148.

Streiner, D.L., Norman, G.R., 2008. *Health Measurement Scales*, 4th edition. Oxford University Press.

Subedar, H., 2005. Nursing Profession: Production of nurses and proposed scope of practice, In: *The South African Health Review*. Health Systems Trust (HST). Pretoria.

Twomey, M., Wallis, L.A., Myers, J.E., 2007. Limitations in validating emergency department triage scales. *Emergency Medicine Journal* 24, 477-479.

van der Wulp, I., van Stel, F., 2009. Adjusted weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *Journal of Clinical Epidemiology* 62, 1196-1201.

Wallis, L.A., 2006. The Cape Triage Score: update. *Emergency Medicine Journal* 23, 740.

Wildschut, A., Mqolozana, T., 2008. Shortage of Nurses in South Africa: Relative or Absolute? Case study report forming part of the HSRC study: A multiple source identification and verification of scarce and critical skills in the South African Labour market. Available at:

<http://www.labour.gov.za/documents/research-documents/shortage-of-nurses-in-south-africa-relative-or-absolute>

Worster, A., Fernandes, C.M., 2004. Paper case scenarios in the assessment of triage tools. *Annals of Emergency Medicine* 44, 524-526.

Worster, A., Gilboy, N., Fernandes, C.M., Eitel D., Eva, K., Geisler, R., Tanabe, P., 2004. Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *Canadian Journal of Emergency Medicine: Journal of Canadian Association of Emergency Physicians* 6, 240-245.

Worster, A., Sardo, A., Eva, K., Fernandes, C.M., Upadhye, S., 2007. Triage Tool Inter-rater Reliability: A Comparison of Live Versus Paper Case Scenarios. *Journal of Emergency Nursing* 33, 319-323.

AGREE for Windows [computer program]. Version 7.002 1999. University of Groningen. Available at [URL:http://www.scienceplus.nl/component/page,shop.product_details/flypage.shop.flypage_splus/product_id,59/category_id,13/manufacture_id,0/option,com_virtuemart/Itemid,26/](http://www.scienceplus.nl/component/page,shop.product_details/flypage.shop.flypage_splus/product_id,59/category_id,13/manufacture_id,0/option,com_virtuemart/Itemid,26/)

Intercooled Stata statistical software package version 9.2 for Macintosh, 21 Nov 2006. Stata Corporation 4905 Lakeway Dr College Station, TX 77845 USA. Available at <http://www.stata.com/>

SPSS statistical software package version 17.0.0, Aug 23, 2008. SPSS Inc., an IBM Company
Headquarters, 233 S. Wacker Drive, 11th floor, Chicago, Illinois 60606. Available at
<http://www.spss.com/>

ADULT TRIAGE SCORE								© South African Triage Group 2008	
	3	2	1	0	1	2	3		
Mobility				Walking	With Help	Stretcher/ Immobile		Mobility	
RR		less than 9		9-14	15-20	21-29	more than 29	RR	
HR		less than 41	41-50	51-100	101-110	111-129	more than 129	HR	
SBP	less than 71	71-80	81-100	101-199		more than 199		SBP	
Temp		Cold OR Under 35		35-38.4		Hot OR Over 38.4		Temp	
AVPU		Confused		Alert	Reacts to Voice	Reacts to Pain	Unresponsive	AVPU	
Trauma				No	Yes			Trauma	
over 12 years / taller than 150cm									
Colour	RED		ORANGE		YELLOW		GREEN		BLUE
TEWS	7 or more		5-6		3-4		0-2		DEAD
Target time to treat	Immediate		less than 10 mins		less than 60 mins		less than 240 mins		DEAD
Mechanism of injury			High energy transfer						
Presentation			Shortness of breath - acute				ALL OTHER PATIENTS		
			Coughing blood						
			Chest pain						
			Haemorrhage - uncontrolled						
	Seizure - current		Seizure - post ictal						
			Focal neurology - acute						
			Level of consciousness reduced						
			Psychosis / Aggression						
			Threatened limb						
	Burn - face / inhalation		Dislocation - other joint		Dislocation - finger or toe				
			Fracture - compound		Fracture - closed				
			Burn over 20%		Burn - other				
			Burn - electrical						
	Burn - circumferential								
	Burn - chemical								
			Poisoning / Overdose		Abdominal pain				
	Hypoglycaemia - glucose less than 3		Diabetic - glucose over 11 & ketonuria		Diabetic - glucose over 17 (no ketonuria)				
			Vomiting - fresh blood		Vomiting - persistent				
Pregnancy & abdominal trauma or pain			Pregnancy & trauma						
			Pregnancy & PV bleed						
Pain			Severe		Moderate		Mild		
Senior Healthcare Professional's Discretion									

Appendix 1: The South African Triage Scale

Vignette Description

1: A fragile looking 29 year old woman is assisted into your ED on crutches. She is breathing very rapidly and says that she has been coughing and vomiting for the past few days. She states that she is not in much pain (2/10). She explains to you that she has not been well enough to look after her child. RR 30, HR 160, BP 109/65, 37.5°C

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

2: "I have had chest pain for 2 weeks now and have a sensation of pins and needles in my left arm and face" reports a 26 year old woman that is carried into your ED on a stretcher. She does not complain of any pain but looks very fearful. RR 23, HR 72 , BP 115/60, 36.3°C, HGT 4.4, HB 11.5

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

3: Paramedics walk in a 48 year old woman who complains of a moderate headache (5/10) and weakness in her left arm. She is mobile and states that she has had a previous CVA. She is tearful and anxious. RR 20, HR 117, BP 205/136, 36°C, HGT 7.9, HB 10

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

4: A 26 year old man is brought into your ED by paramedics. He is on a stretcher. He is screaming and shouting about his severe abdominal pain (9/10). He starts throwing his clothes around the ED and becomes disruptive. RR 20, HR 65, BP 131/95, 36.5°C HGT 5.9 HB 14.5

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

Appendix 2: Acuity assessment – examples of vignettes

Inter-rater reliability of the South African Triage Scale: assessing two different cadres of health care workers in a real-time environment

Paper overview

This paper presents the inter-rater reliability of the South African Triage Scale when used in a real-time environment by both senior health care professionals and ENAs on a general EC patient population at primary care level in South Africa. It also highlights that agreement during triage is higher among nurses that perform the function independently as opposed to performing it as a collaborative doctor-nurse team.

Contribution to the thesis and novelty

This paper addresses the second objective with a focus on comparing inter-rater reliability of individual versus team triage ratings. Contributions of the paper that are unique are that the study was undertaken in a real-time environment at primary care level.

Contribution of candidate

The candidate was responsible for data management and conducted all the analysis in the manuscript. The candidate wrote and managed all drafts of the manuscript, incorporating suggestions from co-authors.

Publication status

The paper was accepted 26 July 2011 for publication in the African Journal of Emergency Medicine.

Motivation for submission to the African Journal of Emergency Medicine

The South African Triage Scale is designed for the African setting and is currently being used in six sub-Saharan countries. This peer-reviewed journal was therefore chosen as the most appropriate platform for disseminating current research findings relating to the reliability of the South African Triage Scale when used in a less developed context.

Title:

Inter-rater reliability of the South African Triage Scale: assessing two different cadres of health care workers in a real-time environment

Authors:

M.Twomey¹, A de Sá¹, LA Wallis², Myers JE¹

Institutions:

- 1) School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
- 2) Division of Emergency Medicine, Department of Surgery, University of Cape Town, Stellenbosch University, Cape Town, South Africa

Word Count

Abstract: 237

Paper: 1554

ABSTRACT

OBJECTIVE: To estimate the inter-rater reliability of triage ratings within individual cadres of health care workers (HCWs) and between different cadres of HCWs using the South African Triage Scale (SATS).

METHODS: Five final year medical students (FMSs), two enrolled nurses (ENs) and two enrolled nursing assistants (ENAs), who had all been trained in the use of the SATS, were selected to prospectively triage Emergency Centre (EC) patients in real time. 25 patients were triaged twice on the first day by individual participants, and another 25 were triaged twice by a collaborative team on the second day. Quadratically weighted kappa (QWK) point estimates were calculated with 95% confidence intervals to assess agreement.

RESULTS: For 25 patients analysed on day one, the QWK values were very high within professions triaging individually: among the FMSs (QWK = 0.94; 95% CI: 0.82-1.0), among the ENs (QWK = 0.92; 95% CI: 0.74-1.0) and moderate between the FMSs and ENs (QWK = 0.57; 95% CI: 0.33-0.81). For 25 patients analysed on day two a team of ENA and FMS, triaging collaboratively, demonstrated moderate agreement (QWK = 0.65; 95% CI: 0.46-0.85).

CONCLUSION: The inter-rater reliability of SATS ratings is excellent within individual HCWs, but significantly lower between different HCWs. This confirms previous reliability studies of the SATS using vignettes and if validated by larger studies would support the feasibility of further implementation of the SATS in primary health care settings across the Western Cape.

Keywords: Emergency Centre, triage, inter-rater reliability, South African Triage Scale

1. INTRODUCTION

The South African Triage Scale (SATS) is an initial measure of patient acuity in the emergency centre (EC); it has been implemented and evaluated since 2006 (1), and has been shown to halve patient mortality (2). The SATS assesses medical urgency based on age appropriate physiologic parameters and a list of clinical discriminators (**Appendix A** shows the adult version) (3). The SATS was intentionally designed for use by Enrolled Nursing Assistants (ENA), due to the limited numbers of doctors and nurses in South Africa (4). ENAs are entry-level nurses who have qualified with a one-year certificate. Previous studies provide evidence of ENA competence using the SATS (1). Final year medical students (FMSs) are also expected to triage. It is part of their training, and task shifting is encouraged especially where patients present straight to the EC with higher acuity levels (5). This introduces a teamwork element, with nurses and FMSs fulfilling several functions as a team, one of them being triage.

A reliability study using vignettes showed that the SATS provides reliable acuity ratings when used by nurses and Emergency Physicians in South Africa (5). However, to date no study has assessed the reliability of South African triage scale ratings of real patients. ECs at Community Health Centres (CHCs) operate 24 hours a day, providing emergency care to the uninsured population; they are staffed by nurses and medical officers, with clinical governance being provided by a Family Physician, and form a key part of the Primary Health Care system. These ECs see the full range of emergency cases, and in Cape Town 24-hour CHCs have been triaging unscheduled patients with SATS since 2006 (6). FMSs rotate through a 24-hour CHC and are required to complete a project as part of their training. Five FMSs were involved in this study as a student project and were included to represent a different cadre of HCW.

We undertook a pilot study to determine the reliability of SATS ratings of enrolled nurses (ENs), ENAs and FMSs on a general undifferentiated patient population at a 24-hour CHC. We aimed to answer the following questions: What is the inter-rater reliability: (i) among nurses triaging individually (ii) among FMSs triaging individually (iii) between nurses and FMSs triaging individually (iv) among the nurse and FMS groups triaging as a collaborative?

2. METHODS

2.1 Study design and data collection

The study was conducted on patients presenting as undifferentiated, unscheduled walk in cases in a 24-hour CHC. A convenience sample of five FMSs and four nurses underwent SATS training during a 3-hour session. For some of the nurses this served as refresher training. After training, the FMSs and nurses (two ENs on day one, and two ENAs on day two) assigned triage acuity ratings to a total of 50 patients over two days. 25 patients were triaged twice on the first day by individual participants while unaware of the ratings given by the other (seven patients received a rating from two different ENs;

eight patients received a rating from two different FMSs; ten patients received a rating from an EN and then an FMS). On the second day another 25 patients were triaged twice collaboratively by two ENA/FMS teams, unaware of the ratings given by the other team. **Figure 1** illustrates the design and how patients were triaged over the two days.

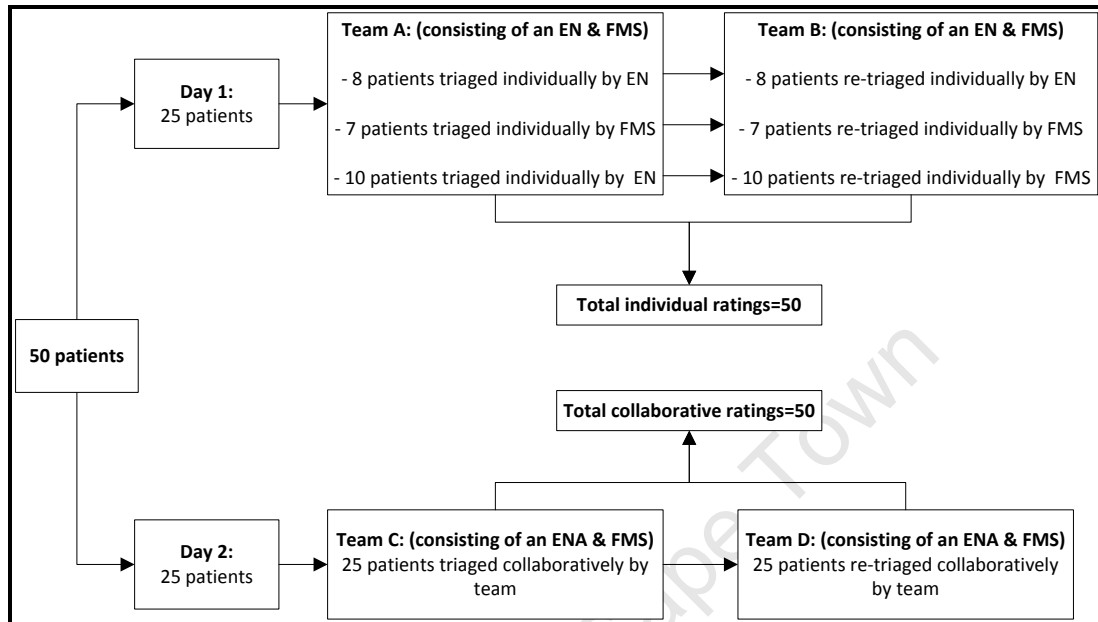


Figure 1: Design and triage process over two days

DAY 1 - one member of Team A individually triaged each patient in Room 1; patients then proceeded to Room 2 where a member of Team B individually re-triaged the same patient. Each member was blinded to the other triage decision and the staff member who had undertaken the triage; no discussion was allowed within or between teams.

DAY 2 – The two ENs and two FMSs from DAY 1 were replaced by two ENAs and two different FMSs. Team C (consisting of an FMS and ENA) collaboratively triaged each patient in Room 1; patients then proceeded to Room 2 where Team D collaboratively re-triaged the same patient. Discussion was allowed within teams on day 2 (i.e. in Team C the ENA could assist the FMS and vice versa), but each team was blinded to the other team's triage decision.

To prevent interference with the care of patients, all patients presenting to Team A or C, and categorized as red or orange, were sent to the resuscitation room immediately and the member of Team B or D was notified and followed the patient to re-triage them there. It is expected that this would have introduced a bias.

2.2 Sampling

A total of 50 patients (representing about 10% of the weekly load of unscheduled patients) were selected via convenience sampling (as they arrived at the CHC) . Patient triage information was captured on a standard form including age, presenting complaint, vital signs and folder number (no other patient details were collected, to ensure anonymity and confidentiality). All nurses and FMSs used the SATS to triage patients into one of four triage categories (Red, Orange, Yellow or Green) (Table 1).

SATS Category	Response	Description of category
RED	Immediate	Emergency
ORANGE	Within 10 min	Very Urgent
YELLOW	Within 60 min	Urgent
GREEN	Within 240 min	Routine
BLUE	Certify	Dead on Arrival

Table 1: South African Triage Scale Categories (3)

2.3 Data analysis

Microsoft Excel (© Microsoft Corporation) was used to capture the data daily on a personal password protected computer. Inter-rater reliability was calculated using the quadratically weighted kappa (QWK) statistic (7). STATA statistical software package version 9.2, 21 Nov 2006 (8) was used to assess inter-rater reliability for the group of nurses and students individually and for the joint group as a team. 95% confidence intervals were calculated for all measures and point estimates were interpreted using the classification model designed by Landis and Koch (9) for the interpretation of kappa.

2.4 Ethical Considerations

Patient care was not delayed as folders were kept in normal administrative flow. Patients initially triaged as red or orange were re-triaged in the resuscitation room. All invasive testing which forms part of the triage process (e.g. blood sugar level, urine dipstick etc.) was only done once (if the team that was re-triaging felt a test was necessary and it had already been performed, the result was made available to them). Verbal consent was obtained from all patients (written consent would have interfered with triage and possibly affected the care of the patient). Only folder numbers were recorded, no names or birthdates were used.

This study was granted ethics approval from the Research Ethics Committee, University of Cape Town (Rec ref 313/2008).

3. RESULTS

12% of patients included in this study were children. **Figure 2** shows the distribution of assigned triage acuity for the 68% of patients (n=34) where there was exact agreement on triage acuity level across the two teams.

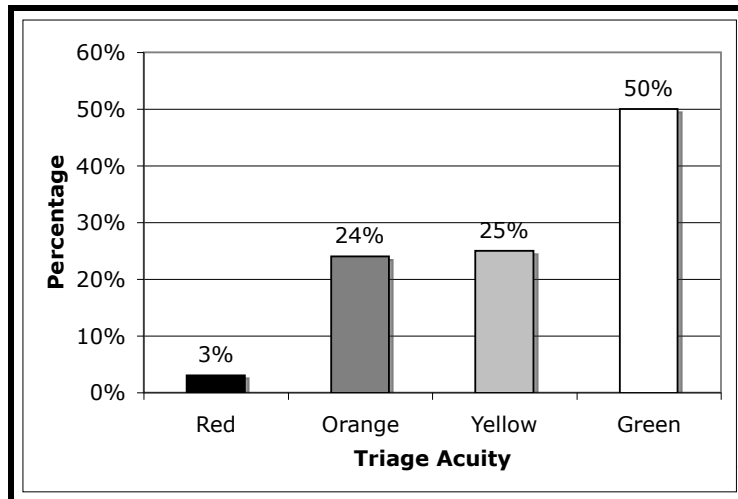


Figure 2: Distribution of triage acuity (n=34)

The inter-rater reliability among ENs and FMSs using the SATS individually was excellent, with a QWK of 0.92; (95% CI: 0.74-1.0) for ENs and QWK of 0.94; (95% CI: 0.82-1.0) for FMSs. Reliability was significantly lower between individual FMSs and ENs (QWK=0.57; 95% CI: 0.33-0.81). When triaging as a collaborative team the QWK was 0.65; (95% CI: 0.46-0.85).

4. DISCUSSION

This pilot study evaluated the reliability of individual versus team triage SATS ratings when used by nurses and FMSs on real patients at primary care level. Our results revealed excellent inter-rater reliability of acuity ratings within separate professions (among ENs and among FMSs) using the SATS individually. This dropped significantly when inter-rater reliability was assessed between ENs and FMSs, and also dropped for collaborative team triage.

Individual triage ratings across professions showed only moderate reliability. This may be attributed to the tendency of final year medical students beginning the process of differential diagnosis during triage and generally 'downgrading' patients as further assessments are done. Anecdotally it was found that students tended to take a longer more comprehensive history and 'overrule' triage discriminators when the likely diagnosis was not in exact accordance with the discriminator e.g. A person with chest pain according to SATS is orange (in order for ischemic heart disease to be ruled out urgently). The SATS has not been designed to provide a diagnosis, but rather guide the triage practitioner through a stepwise approach to safely sort patients according to medical urgency.

Further investigations are required to determine whether this may be the reason for lower inter-rater reliability between different professions.

5. LIMITATIONS

The principal limitation of this study is the small sample size, along with potential bias from sending patients initially triaged as high priority to the resuscitation room, to be retriaged there.

6. CONCLUSION

This pilot study shows that inter-rater reliability within individual cadres of HCWs using the SATS is excellent. Inter-rater reliability was significantly reduced between different cadres of HCWs using the SATS. This confirms previous findings of SATS reliability studies in South African ECs (5). Future studies should include more sites, increase the sample size and utilize qualitative methods to explore the thought processes of different professions at triage.

ACKNOWLEDGEMENTS

The authors wish to thank the final year medical students (Azra Ghoor, Joel Giddey, Daniel Giles, John-D Lotz and Aa-i-shah Manan) involved in this project, as well as the nurses and staff at Retreat CHC for their participation and support.

8. COMPETING INTERESTS

None

9. FUNDING

None

10. REFERENCES

- (1) Bruijns SR, Wallis LA, Burch VC. A prospective evaluation of the Cape Triage Score in the emergency department of an urban public hospital in South Africa. *Emerg Med J* 2008; **25**:398-402.
- (2) Bateman C. New triage system halves mortalities. *SAMJ* 2006; **96**:770–772.
- (3) Gottschalk S, Wood D, DeVries S, Wallis L, Bruijns S, On behalf of the Cape Triage Group . The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006 **23**:149-153.
- (4) Subedar H. Personal Interview at the South African Nursing Council. Pretoria. November 2006.
- (5) Twomey M, Wallis LA, Thompson ML, Myers JE. The South African Triage Scale (Adult Version) provides reliable acuity ratings: A graphical approach enhances conventional methods. Unpublished.
- (6) Policy for the implementation of the triage of patients in Western Cape emergency centres. Policy circular H7 of 2006.
- (7) Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. *Scand J Caring Sci* 2005 **19**:427-431.
- (8) Intercooled Stata statistical software package version 9.2 for Macintosh, 21 Nov 2006. Stata Corporation 4905 Lakeway Dr College Station, TX 77845 USA. Available at <http://www.stata.com/>
- (9) Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977 **33**:159-174.

ADULT TRIAGE SCORE								© South African Triage Group 2008	
	3	2	1	0	1	2	3		
Mobility				Walking	With Help	Stretcher/ Immobile		Mobility	
RR		less than 9		9-14	15-20	21-29	more than 29	RR	
HR		less than 41	41-50	51-100	101-110	111-129	more than 129	HR	
SBP	less than 71	71-80	81-100	101-199		more than 199		SBP	
Temp		Cold OR Under 35		35-38.4		Hot OR Over 38.4		Temp	
AVPU		Confused		Alert	Reacts to Voice	Reacts to Pain	Unresponsive	AVPU	
Trauma				No	Yes			Trauma	
over 12 years / taller than 150cm									
Colour	RED	ORANGE	YELLOW	GREEN	BLUE				
TEWS	7 or more	5-6	3-4	0-2	DEAD				
Target time to treat	Immediate	less than 10 mins	less than 60 mins	less than 240 mins	ALL OTHER PATIENTS				
Mechanism of injury		High energy transfer							
Presentation		Shortness of breath - acute							
		Coughing blood							
		Chest pain							
		Haemorrhage - uncontrolled		Haemorrhage - controlled					
		Seizure - current	Seizure - post ictal						
		Focal neurology - acute							
		Level of consciousness reduced							
		Psychosis / Aggression							
		Threatened limb							
		Dislocation - other joint		Dislocation - finger or toe					
	Burn - face / inhalation	Fracture - compound		Fracture - closed					
		Burn over 20%	Burn - other						
		Burn - electrical							
		Burn - circumferential							
	Burn - chemical								
		Poisoning / Overdose	Abdominal pain						
Hypoglycaemia - glucose less than 3	Diabetic - glucose over 11 & ketonuria	Diabetic - glucose over 17 (no ketonuria)							
	Vomiting - fresh blood	Vomiting - persistent							
	Pregnancy & abdominal trauma or pain	Pregnancy & trauma							
		Pregnancy & PV bleed							
Pain		Severe	Moderate	Mild					
Senior Healthcare Professional's Discretion									

Appendix A: The South African Triage Scale

Limitations in validating emergency department triage scales

Paper overview

This paper explores methods of emergency centre triage scale validation and considers their application in developing countries. It discusses the limitations inherent in validating triage scales, and suggests a new consensus building validation approach as an alternative for validating triage scales in less developed countries.

Contribution to the thesis and novelty

This paper addresses the second objective. There is currently no clarity in the emergency medicine literature on the most appropriate method for validating triage scales. Other contributions of the paper that are unique are the discussion of limitations in validating triage scales with an emphasis on less developed countries.

Contribution of candidate

The candidate wrote and managed all drafts of the manuscript, incorporating suggestions from co-authors.

Publication status

Published in April 2007



Limitations in validating emergency department triage scales

Michele Twomey, Lee A Wallis and Jonathan E Myers

Emerg. Med. J. 2007;24:477-479
doi:10.1136/emj.2007.046383

Updated information and services can be found at:
<http://emj.bmj.com/cgi/content/full/24/7/477>

These include:

References

This article cites 16 articles, 5 of which can be accessed free at:
<http://emj.bmj.com/cgi/content/full/24/7/477#BIBL>

Rapid responses

One rapid response has been posted to this article, which you can access for free at:
<http://emj.bmj.com/cgi/content/full/24/7/477#responses>

You can respond to this article at:
<http://emj.bmj.com/cgi/eletter-submit/24/7/477>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article

Topic collections

Articles on similar topics can be found in the following collections
[Other accident and emergency medicine](#) (1864 articles)

Notes

To order reprints of this article go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to *Emergency Medicine Journal* go to:
<http://journals.bmj.com/subscriptions/>

ORIGINAL ARTICLE

Limitations in validating emergency department triage scales

Michele Twomey, Lee A Wallis, Jonathan E Myers

Emerg Med J 2007;24:477-479. doi: 10.1136/emj.2007.046383

See end of article for authors' affiliations

Correspondence to: Michele Twomey, School of Public Health, University of Cape Town, Cape Town, South Africa; satriage@webmail.co.za

Accepted 29 April 2007

Objective: To examine whether current validation methods of emergency department triage scales actually assess the instrument's validity.

Methods: Optimal methods of emergency department triage scale validation are examined in developed countries and their application to developing countries is considered.

Results and conclusion: Numerous limitations are embedded in the process of validating triage scales. Methods of triage scale validation in developed countries may not be appropriate and repeatable in developing countries. Even in developed countries there are problems in conceptualising validation methods. A new consensus building validation approach has been constructed and recommended for a developing country setting. The Delphi method, a consensual validation process, is advanced as a more appropriate alternative for validating triage scales in developing countries.

Emergency department (ED) triage is the process of sorting and filtering patients based on medical priority. It aims to determine a patient's acuity level in order to facilitate timely and effective care before their condition worsens. A patient's acuity level is defined as the urgency for effective care. In the ED triage setting effective care is defined as the provision of an intervention or treatment that reduces the patient's urgency for care or prevents clinical deterioration.¹ If patients receive timely and effective care, triage has achieved its purpose (as seen at point A in fig 1).

This illustration of triage is a highly simplified approach to a complex set of interrelationships. It is acknowledged that additional variables may influence optimal time to care and effectiveness of care significantly (such as variability in triage nurse decisions).

RELIABILITY

The evaluation of a triage tool involves assessing reliability and validity.² Reliability refers to the degree to which repeated

assessments of the same patient with a triage instrument will deliver the same acuity level. *Inter-rater* reliability determines whether there is significant variability between different triage officers rating the same patient, and *intra-rater* reliability assesses the variability within a single triage officer re-rating the same patient. Reliability makes no reference to a criterion, and so only illustrates consistency with triage repetition. It shows nothing about its validity (whether it is a reflection of the truth). A measure can therefore be highly reliable without being valid.³

Reliability can be estimated by evaluating different types of agreement. Percentage agreement, the κ coefficient and the weighted κ coefficient are three common ways of measuring agreement between raters,⁴ but these measures can generate quite different values. Measuring only the percentage agreement is not recommended because it does not take into account agreement expected on chance alone.⁵ The κ coefficient considers both percentage agreement between raters and percentage agreement expected by chance; unfortunately, it does not take into account the magnitude of disagreement, which may become significant in ordinal data. As a result, the weighted κ coefficient has become the instrument of choice as it assigns different weights of agreement according to the magnitude of disagreement, and enables more explicit comparisons between different studies.⁴ While the majority of research in triage has focused on *inter-rater* and *intra-rater* reliability, which has its uses, it is of greater importance to determine whether a triage tool is in fact valid. We will therefore be focusing on the validity of a triage tool rather than its reliability.

VALIDITY

Validity refers to the degree with which the measured acuity level reflects the patient's true acuity at the time of triage. The term valid implies that there is some sort of external reference or "gold standard" which by definition has absolute accuracy.³ Studies that aim to see how closely an instrument approximates the truth, test *criterion validity*. Unfortunately it is not possible to measure the truth for patient acuity,⁶ as there are myriad events that can occur from the time that a patient presents to the ED to the time of discharge (including the length of time to initiation of care, the quality of that care, and

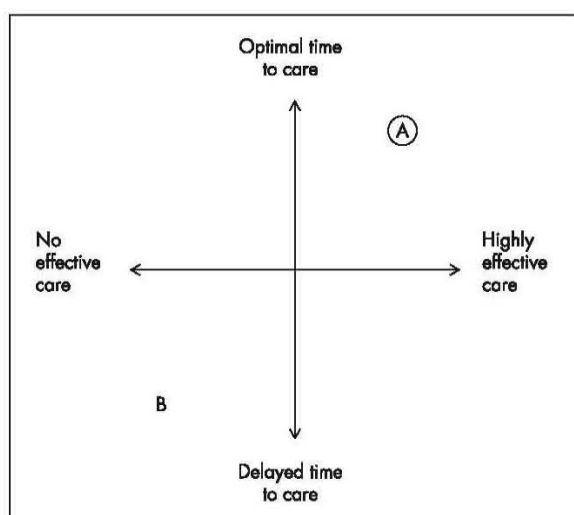


Figure 1 Triage scales are most valuable at (A), where triage facilitates optimal time to highly effective care, and least valuable at (B), where triage facilitates delayed time and ineffective care.

Abbreviations: ATS, Australasian Triage Scale; CTAS, Canadian Triage Acuity Scale; ED, emergency department; ESI, Emergency Severity Index; ETAT, Emergency Triage Assessment and Treatment; MEWS, Modified Early Warning Score

www.emjonline.com

non-medical factors influencing disposal—for example, social factors). As a result, surrogate outcome markers have been used as criteria to assess validity. This has led to other ways of assessing validity for ED triage tools. The two most commonly found in the literature are tests of predictive or consensual validity. These have been approached in a unifying manner by Streiner and Norman, who reconceptualise a variety of notions of validity commonly used in the literature as *construct validity*.⁷

There is a hierarchy of validity testing in which criterion is the best (table 1). Streiner and Norman have shown that unlike the traditional classification of validity, predictive, consensual and other types of validity are all seen as variants of construct validity.⁷ Typically in developed countries, criterion validity methods are used.

We will use Streiner and Norman's conceptual framework to answer the following questions:

- Do current methods of triage tool validation actually assess the validity and what are the limitations underlying these methods?
- How can these limitations be overcome with special reference to developing countries?

CURRENT METHODS OF TRIAGE TOOL ASSESSMENT AND THEIR LIMITATIONS

A number of different triage systems are used in developed countries. To date, four reliable ordinal ED triage scales have been researched and published: the Australasian Triage Scale (ATS),⁸ the Canadian Triage Acuity Scale (CTAS),⁹ the Emergency Triage Scale (aka Manchester Triage Scale)¹⁰ and the Emergency Severity Index (ESI).¹¹ While there has been some focus on the reliability of triage tools, not much is published on their validity. *Predictive validity* (a type of construct validity) is the most frequently used method of assessing tools.¹² It considers the degree to which the triage acuity level is able to predict true acuity. Particular outcomes, or events with time-ordering, are selected as surrogate markers (such as mortality rates, hospital admission rates, resource utilisation, and length of stay in hospital). There are methodological problems with the use of this type of validity as it does not always answer the core question: "Is the triage instrument able to measure what is supposed to be measured?" In patients it does not measure acuity at the time of assessment (and is inherently confounded by the effectiveness of the health care intervention).

Examples of predictive validity abound in the triage literature, as surrogate outcome markers are practical to measure and are claimed to be closely associated with true acuity.³ This has compelled clinicians and researchers to utilise triage instruments as prediction tools. However, our ability to identify and measure the relationship between patient acuity level and outcome depends not only on the measurement of the surrogate outcome marker and the patient's acuity level, but also very importantly on confounding variables such as variability in triage nurse decisions, and delayed and ineffective treatment. These may affect the surrogate outcome marker.

Table 1 Traditional validity testing versus Streiner and Norman's framework

Traditional	Streiner and Norman
Criterion	Criterion
Construct	Construct
Predictive	Construct
Consensual	Construct

HOW CAN THESE LIMITATIONS BE OVERCOME?

A detailed literature review revealed that very little has been published on triage in developing countries. The World Health Organization reports that triage research is not a priority in low- to middle-income countries.¹³ They have accordingly developed the Emergency Triage Assessment and Treatment (ETAT)¹⁴ for application to developing countries. While this subjective system has been successfully implemented in Malawi, countries like India, Brazil and South Africa have sought a more objective triage instrument based on physiology. They have either adopted the triage instrument from a developed country or modified it to their own local context and needs (Patriacia Neto, Quinta D'or Hospital, Rio de Janeiro, May 2007, personal communication). South Africa has adapted the Modified Early Warning Score (MEWS) as the South African Triage Scale after validating it on the local national population.¹⁵ Some areas of Brazil have adopted the CTAS, others the ESI.

During any validity testing an important distinction needs to be made between internal validity (which refers to inferences about the source population), and external validity (whether inferences may be generalised to people outside the source population).¹⁶ A triage tool designed for a developed country may be valid in that context, leading to favourable results that are meaningful and have implications for action. If, however, the same triage tools were applied in a developing country, results may vary due to different resources and skills. Similarly results may vary when applying surrogate markers from developed countries to undertake validity testing in developing countries. This variability may increase the random error in both triage acuity level and outcome category; it would therefore be more appropriate to apply a locally developed tool that is meaningful in the local context (has internal validity), but that may not be applicable in a developed country (lack of external validity).

Whichever tool is used, an assessment of its usefulness in these settings is required. When selecting surrogate outcome markers (such as mortality rates, hospital admission rates, resource utilisation, and length of stay in hospital), it is assumed that there is systematic record keeping, and that the care given is effective. While this may often be the case in developed countries, it is typically not the case in developing countries. Poor record keeping and ineffective care may have significant effects on surrogate outcome markers and patients' final dispositions. Markers such as these are imperfect measures of patient acuity in the developing world. It is thus important to identify and measure all confounding variables that may be affecting the surrogate outcome marker: given the poor record keeping and lack of efficiency, this is unlikely to be feasible in developing countries.

Delphi methodology

The Delphi method was developed in the 1950s by the RAND Corporation in California, USA.¹⁷ The technique has diversified and is being applied to more mainstream social sciences, in business and, in the last two decades, within the healthcare arena.¹⁸ It is a consensus building technique designed to gain insight into a particular field to enable decision making in areas where published information is inadequate or non-existent.¹⁹ The approach of the Delphi technique is to establish a panel of appropriate experts that have agreed to complete an iterative process on a particular issue, with the key objective being to reach consensus.²⁰ Panelist anonymity is maintained throughout the process and controlled feedback is provided from each iterative round, resulting in a statistical aggregation of the group response.¹⁸

The Delphi method is another form of construct validity that may be useful when assessing triage scales in developing countries. It allows the development of a surrogate "gold standard" determined by specialist panel consensus. The triage

tool's validity may then be tested against this construct of true underlying acuity that is consensually arrived at. There appear to be only very few examples in the world literature that elaborate on the use of this form of construct validity.

Wallis *et al*²¹ used consensus from Delphi methodology to establish triage acuity levels against which to test pre-hospital mass casualty triage tools: such methodology may be used in ED triage tool assessment.

There are several reasons why the Delphi methodology is best suited to assessing ED triage tools in developing countries. The Delphi technique eliminates potential bias due to individual group dynamics and is financially feasible.

Limitations of the Delphi technique are mostly a result of poorly conducted studies rather than fundamental problems. One of the weaknesses cited is that the response rates can be low and often decrease as the rounds progress. However, non-response is typically very low in practice, since most researchers have personally obtained assurance of participation. Similarly attrition tends to be low and the researcher can easily ascertain the cause by talking with the dropouts.²² Selection of the Delphi panel depends on the research question. Problems may arise with a lack of representativeness in that only experts with an interest and involvement will become participants. Another potential weakness of the Delphi as a consensus method is that it overlooks important minority issues because it tries to obtain consensus.²³ However, despite these limitations we believe that the Delphi process is the most appropriate form with which to test the validity of triage tools in the developing world.

CONCLUSION

In developing countries a form of construct validity derived from a consensual process appears to be the most appropriate form of validation of triage tools. This is due to lack of criteria for true acuity, confounding variables that relate to differential health care resources by level of development, and lack of external validity of other triage scales.²⁴ We propose the Delphi method when testing the South African Triage Scale. This is an example of construct validity testing in the developing world.

Authors' affiliations

Michele Twomey, School of Public Health, University of Cape Town, Cape Town, South Africa

Lee A Wallis, Division of Emergency Medicine, University of Cape Town & Stellenbosch University, Capetown, South Africa

Jonathan E Myers, Occupational and Environmental Health Research Unit, University of Cape Town, Capetown, South Africa

Funding: None

Competing interests: None

Contributions: LW had the original idea; MT wrote the first draft; both authors contributed to the final article.

REFERENCES

- 1 Bindman AB. Triage in accident and emergency departments. *BMJ* 1995;311:404.
- 2 Rutschmann OT, Kossovsky M, Geissbühler A, *et al*. Interactive triage simulator revealed important variability in both process and outcome of emergency triage. *J Clin Epidemiol* 2006;59:615–21.
- 3 Lewis RJ. Reliability and validity: meaning and measurement. Presented at 1999 Annual Meeting of the Society for Academic Emergency Medicine (SAEM) in Boston, Massachusetts. <http://www.ambpeds.org/ReliabilityandValidity.pdf>.
- 4 Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. *Scand J Caring Sci* 2005;19:427–31.
- 5 Fernandes CB, Groth SJ, Johnson LA, *et al*. A uniform triage scale in emergency medicine. American College of Emergency Physicians, 1999. <http://www.acep.org/NR/rdonlyres/EE2B10F7-51BE-42A8-94B7-779726144017/0/triagescaleip.pdf>.
- 6 Fernandes CM, Tanabe P, Gilboy N, *et al*. Five-level triage: a report from the ACEP/ENA five-level triage task force. *J Emerg Nurs* 2005;31:39–50.
- 7 Streiner DL, Norman GR. "Precision" and "accuracy": Two terms that are neither. *J Clin Epidemiol* 2006;59:327–30.
- 8 Australasian College of Emergency Medicine. Guidelines for the Implementation of the Australasian Triage Scale in Emergency Departments. Australasian College of Emergency Medicine, 1998.
- 9 Canadian Association of Emergency Physicians. Implementation guidelines for the Canadian Emergency Department Triage and Acuity Scale (CTAS). Canadian Association of Emergency Physicians, 1998.
- 10 Manchester Triage Group. *Emergency triage*. London: BMJ Publishing Group, 1997.
- 11 Gilboy N, Tanabe P, Travers DA, *et al*. The Emergency Severity Index implementation handbook: a five-level triage system. Des Plaines, Illinois: Emergency Nurses Association, 2003.
- 12 Fan J, Al Darab A, Eva K, *et al*. Triage scales in the emergency department: a systematic review. *Ann Emerg Med* 2005;46:S41.
- 13 Razzak, Junaid A, Kelleman, Arthur L. Emergency medical care in developing countries: is it worthwhile? *Bull World Health Organ* 2002;80:900–5.
- 14 Robertson MA, Molyneux EM. Triage in the developing world – can it be done? *Arch Dis Child* 2001;85:208–13.
- 15 Gottschalk SB, Wood D, de Vries S, *et al*. The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006;23:149–53.
- 16 Rothman KJ, Greenland S. *Modern epidemiology*. Philadelphia: Lippincott-Raven Publishers, 1998:115–47.
- 17 Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science* 1963;9:458.
- 18 Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995;311:376–80.
- 19 Pill J. The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio-Econ Plan Sci* 1971;5:57–71.
- 20 Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–15.
- 21 Wallis LA, Carley S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. *Emerg Med J* 2006;23:291–5.
- 22 Okali C, Pawlowski S. The Delphi method as a research tool: an example, design considerations and applications. *Information and Management* 2004;42:15–29.
- 23 Van Teijlingen E, Pitchforth E, Bishop C, *et al*. Delphi method and nominal group techniques in family planning and reproductive health research. *J Fam Plann Reprod Health Care* 2006;32:249252.
- 24 Fernandes CB, Groth SJ, Johnson LA, *et al*. A uniform triage scale in emergency medicine. American College of Emergency Physicians, 1999. <http://www.acep.org/NR/rdonlyres/EE2B10F7-51BE-42A8-94B7-779726144017/0/triagescaleip.pdf>.

The South African Triage Scale (Adult Version) provides valid acuity ratings when used by doctors and enrolled nursing assistants

Paper overview

This paper demonstrates that the South African Triage Scale when used by both doctors and nurses on a general EC patient population provides valid triage acuity ratings.

Contribution to the thesis and novelty

This paper addresses the fourth and fifth objective and one of the important components of any triage scale, namely validity. Contributions of the paper that are unique are the presentation of predictive values as well as the introduction of mis-prediction terms associated with predictive values, over- and under-prediction.

Contribution of candidate

The candidate wrote and managed all drafts of the manuscript, incorporating suggestions from co-authors.

Publication status

The paper has been accepted 26 July 2011 for publication in the African Journal of Emergency Medicine.

Motivation for submission to the African Journal of Emergency Medicine

The South African Triage Scale is designed for the African setting and is currently being used in six sub-Saharan countries. This peer-reviewed journal was chosen as the most appropriate platform for sharing current research findings relating to the validity of the South African Triage Scale when used in a less developed context.

Title:

The South African Triage Scale (Adult Version) provides valid acuity ratings when used by doctors and enrolled nursing assistants

Authors:

M.Twomey¹, LA. Wallis², M. Thompson³, JE. Myers¹

Institutions:

- 1) School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
- 2) Division of Emergency Medicine, Department of Surgery, University of Cape Town & Stellenbosch University, Cape Town, South Africa
- 3) Department of Biostatistics, University of Washington, Seattle, United States

Word Count

Abstract: 239

Paper: 3244

ABSTRACT

OBJECTIVE: To estimate the validity of triage ratings by South African nurses and doctors with training and practical experience using the South African Triage Scale.

METHODS: Five emergency physicians and 10 enrolled nursing assistants, who had been trained in the use of the South African Triage Scale, were selected via convenience sampling to retrospectively triage adult emergency centre vignettes. Participants independently assigned triage ratings to 100 written vignettes unaware of the ratings given by others. Triage ratings were compared with ratings of two experts from the South African Triage Group. Standard validity indicators including sensitivity, specificity, positive predictive value and negative predictive value were used to estimate the validity for the combined group of emergency physicians and enrolled nursing assistants. Associated percentages for over-/under-triage were used to further assess validity within the South African context and over-/under-prediction to further assess practical application of the South African Triage Scale.

RESULTS: On average over all acuity levels, sensitivity was 75%, specificity 91%, under-triage occurred 10% and over-triage 15% of the time. The positive predictive value was 74% and negative predictive value 91%.

CONCLUSION: The results of this study fall within the accepted range of over-/under-triage and indicate that the South African Triage Scale is valid when used by emergency physicians and nurses to triage emergency centre vignettes under South African conditions. Further research into appropriate reference ranges for extent of over-/under-triage and over-/under-prediction within each acuity level is recommended.

1. INTRODUCTION

Emergency centre (EC) triage is the process of sorting and filtering patients based on medical urgency. It aims to determine a patient's acuity level in order to facilitate timely and effective care before their condition worsens. A patient's acuity level is defined as the urgency for effective care. The South African Triage Scale (SATS) (1) is an initial measure of patient acuity in the EC that was developed by the South African Triage Group (SATG) (2). It assesses medical urgency based on physiologic parameters and a list of clinical discriminators (**Appendix 1**). Similar to the Manchester Triage System (MTS) (3) and the Canadian Emergency Department Triage and Acuity Scale (CTAS) (4) the SATS incorporates target times to treatment. Patients are categorised into one of four acuity levels: Red (emergency – should be seen immediately), Orange (very urgent – should be seen in less than 10 minutes), Yellow (urgent – should be seen in less than 60 minutes), Green (routine – should be seen in less than four hours) (5). The SATS was intentionally designed for use by an Enrolled Nursing Assistant (ENA) due to the limited numbers of doctors and professional nurses in South Africa (6). ENAs are entry-level nurses that have qualified with a one-year certificate (7). Previous studies provide evidence of ENA competence and reliability using the SATS (8) (9). The SATS has been implemented, monitored and refined in the public and private health care setting since 2006 (10) (11). To date no study has assessed the validity of the SATS in South African ECs using South African experts as a reference.

The validity of a triage scale is an important measure that tells us how close an acuity rating assigned using that scale is to the true acuity of that patient (12). Reliability is an equally important measure, but it refers to agreement between raters and within raters, using the scale, without reference to the patient's true acuity (12). Previous reports describe the fact that triage scales have no uniquely defined reference standard (1)(13)(14) and therefore one of the challenges in estimating validity lies in the task of meaningfully comparing validity assessments of triage scales across studies and contexts (1). Type of reference standard used (whether it be patient disposition/ length of stay in hospital/ resource utilisation etc. (15-21) or an expert panel (22-27)) will influence the performance characteristics of that scale. Even though triage scales inherently differ depending on their context and design, there should ideally be some uniformity on the most appropriate performance characteristics when reporting on triage scale validity with accepted reference ranges that serve as a guideline and reference criteria.

In the current literature mis-triage is defined as the extent of over-/under-triage relative to true acuity (28). In this study we have classified mis-triage into two different types (i) mis-triage with reference to true acuity levels and (ii) mis-prediction with reference to raters' assigned acuity levels. We refer to mis-triage in our sensitivity analysis as over-/under-triage, and mis-prediction in our analysis of positive predictive values (PPV) as over-/under-prediction.

The objectives of this study are to estimate the validity of the SATS used by both emergency physicians and ENAs on a general adult EC patient population in South Africa. Comparing emergency physician and ENA ratings to those of a local expert panel, we will answer the following questions:

- (1) What is the sensitivity and specificity of the SATS and the associated percentage of over-/under-triage?
- (2) What is the PPV and negative predictive value (NPV) of the SATS and the associated percentage of over-/under-prediction?

2. METHODS

2.1 Study design

A validation study conducted on a series of vignettes that had been collected prospectively from real patients.

2.2 Study sample

Five emergency physicians and ten ENAs were selected using convenience sampling, and invited to participate in our validation study. Individuals came from different geographically located health facilities and represented different sub districts within the Western Cape Province. Validity was assessed using adult vignettes as suitable proxies for live triage cases (29). Based on extensive use in other studies and their advantage in terms of cost and time (30) (31), this method is particularly useful for a less developed country such as South Africa.

100 adult vignettes were prospectively abstracted from randomly selected actual EC case presentations at a secondary hospital and have been previously referenced in a reliability study (8). Vignettes covered characteristics such as gender, age, presenting complaint, mode of arrival and vital signs. **Appendix 2** shows examples.

2.3 Methods of measurement and data collection

ENAs and emergency physicians attending mandatory SATS training sessions in 2009 were required to complete retrospective triage on these vignettes as part of an evaluation exercise. This was done in a classroom environment where candidates were asked to use the SATS to independently triage written sets of vignettes into one of four triage categories (**Table 1**). Five emergency physicians and ten ENAs completed the sets of vignettes. In total 15 raters assessed 100 vignettes (1500 assessments).

SATS Category	Response	Description of category
RED	Immediate	Emergency
ORANGE	Within 10 min	Very Urgent
YELLOW	Within 60 min	Urgent
GREEN	Within 240 min	Routine

Table 1: South African Triage Scale Categories (5)

The SATS categories assigned by the emergency physicians and ENAs were compared to a reference standard that was generated using two local experts from the SATG who had in-depth knowledge of the SATS and experience in its use and application. They independently reviewed the 100 vignettes (with additional information on use of resources, length of stay in hospital and disposal), which allowed for comprehensive expert judgement in generating a reference standard and classified them into an acuity level. The experts' ratings were in perfect agreement for all 100 vignettes. The ratings for each vignette could therefore be aggregated into a single set of "true" acuity levels that served as a reference standard in this study.

For the purposes of this study we will, for clarity, define in **Tables 2 & 3** the measures used to report on validity of the SATS. These measures apply to every triage acuity category and **Table 3** uses the yellow/urgent triage category as an example to elaborate.

		Experts				
		Emergency	Very Urgent	Urgent	Routine	
Raters	Emergency	A	B	C	D	Q
	Very Urgent	E	F	G	H	R
	Urgent	I	J	K	L	S
	Routine	M	N	O	P	T
		U	V	W	Z	

Table 2: A hypothetical contingency table illustrating numbers of triage ratings by experts and raters

sensitivity (%)	$\frac{K}{W}$	vignettes triaged as urgent by the raters amongst all urgent vignettes (as judged by the experts)
specificity (%)	$\frac{(Q-C)+(R-G)+(T-O)}{U+V+Z}$	vignettes triaged as non-urgent by the raters amongst all non-urgent vignettes (as judged by experts)
under-triage (%)	$\frac{O}{W}$	urgent vignettes (as judged by experts) that received a less urgent acuity rating from the raters
over-triage (%)	$\frac{C+G}{W}$	urgent vignettes (as judged by the expert panel) that received a more urgent acuity rating from the raters
Positive Predictive value	$\frac{K}{S}$	vignettes identified as urgent by the experts amongst all triaged as urgent by the raters
Negative Predictive value	$\frac{(U-I)+(V-J)+(Z-L)}{Q+R+T}$	vignettes identified as non-urgent by the experts amongst all triaged as non-urgent by the raters
under-prediction	$\frac{I+J}{S}$	vignettes triaged as urgent by raters which are truly a higher acuity level (as judged by experts)
over-prediction	$\frac{L}{S}$	vignettes triaged as urgent by raters which are truly a lower acuity level (as judged by experts)

Table 3: Example of definitions of performance indicators for the acuity “urgent”

2.4 Data analysis

Validity was assessed by calculating the sensitivity, specificity, and associated over-/under-triage relative to the experts’ acuity assignments; and PPV, NPV, and associated over-/under-prediction relative to the raters’ acuity assignments. Histograms were designed to illustrate and visually compare mis-triage and mis-prediction at each acuity level. Mis-triage was interpreted using the accepted range for average under-triage of not more than 5%-10%, which the American College of Surgeons Committee on Trauma (ACSCOT) considers unavoidable and an associated average over-triage rate of 30%-50% (32). To our knowledge no accepted norms exist for over-/under-triage at each acuity level and no ACSCOT reference ranges exist for predictive values. We therefore used the ACSCOT ranges to interpret the extent of average over-/under-triage only.

The literature indicates that AGREE 7 for Windows (33) is the only software programme that allows calculations of the kappa statistic in relation to a reference standard as well as an option to determine unique weights (28). We found the AGREE 7 package difficult to use, with limited documentation and producing in some cases erroneous results. After several unsuccessful attempts to contact their support service, we chose not to report the kappa statistic and rather focused on the above-mentioned standard validity indicators.

It is known that the kappa statistic depends on the distribution of cases and number of categories, which limits generalizability to settings with different distributions (34-35). In addition the kappa coefficient does not reflect differences in agreement at individual ordinal values (here acuity levels), and therefore only provides a one-dimensional overview (34 36).

3. RESULTS

Five emergency physicians and 10 ENAs each evaluated the 100 vignettes (1500 assessments). **Table 4** summarizes the sensitivity analysis and **Table 5** the predictive value analysis. **Table 4** shows that, on average, under-triage (10%) occurs less frequently than over-triage (15%), relative to the true acuity assigned by experts. **Table 5** shows that, relative to the acuity assigned by the raters, under-prediction (11.4%) occurs on average less frequently than over-prediction (14.8%).

		Vignettes (n)	Triage ratings (n)	Emergency Physicians and ENAs Combined 15 raters for 100 vignettes (a total of 1500 ratings)				Performance relative to experts as the criterion/reference standard			
				Emergency	Very urgent	Urgent	Routine	Sensitivity	Specificity	Over-triage	Under-triage
Expert SATS Category	Emergency	8	120	78.3 *	21.7 z	0.0	0.0	78.3	96.8	n/a	21.7
	Very urgent	45	675	5.8 z	80.9 *	9.5 z	3.8	80.9	82.2	5.8	13.4
	Urgent	33	495	0.8	20.6 z	72.3 *	6.3 z	72.3	89.0	21.4	6.3
	Routine	14	210	0.5	9.0	22.4 z	68.1 *	68.1	95.6	31.9	n/a
Mean								74.9	90.9	14.8	10.3

Table 4: Summary of combined emergency physician and nurse agreement with the experts SATS rating of vignettes (%)

Raters SATS category	Triage ratings (n)	PPV (%)	NPV (%)	Over-prediction (%)	Under-prediction (%)
Emergency	138	68.1	98.1	31.9	n/a
Very urgent	693	78.8	84.4	17.4	3.8
Urgent	469	76.3	86.7	10	13.7
Routine	200	71.5	94.9	n/a	28.5
Mean		73.7	91	14.8	11.5

Table 5: Summary of expert's SATS category agreement with combined emergency physician and nurse ratings for all vignettes

Figure 1 summarizes all vignettes with acuity levels as assigned by the local experts, and illustrates the probability that blinded raters using the SATS will over-/under-triage vignettes at each acuity

level. It shows, for instance, that 22% of the true “emergency” vignettes were under-triaged by one acuity level and that no true “emergency” vignettes were mis-triaged as “urgent” or “routine”.

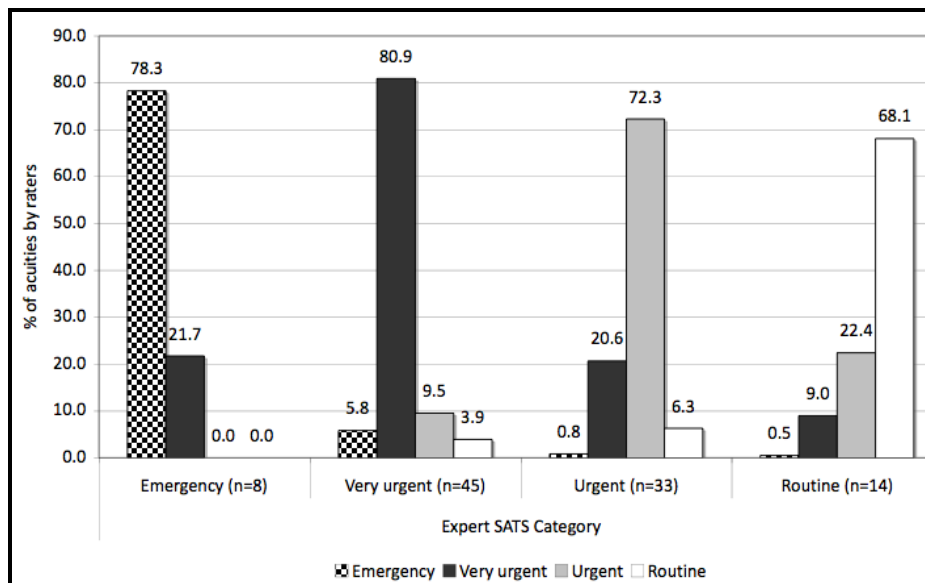


Figure 1: Graphical displays of the sensitivity, over-/under-triage of SATS for emergency, very urgent, urgent and routine acuity levels.

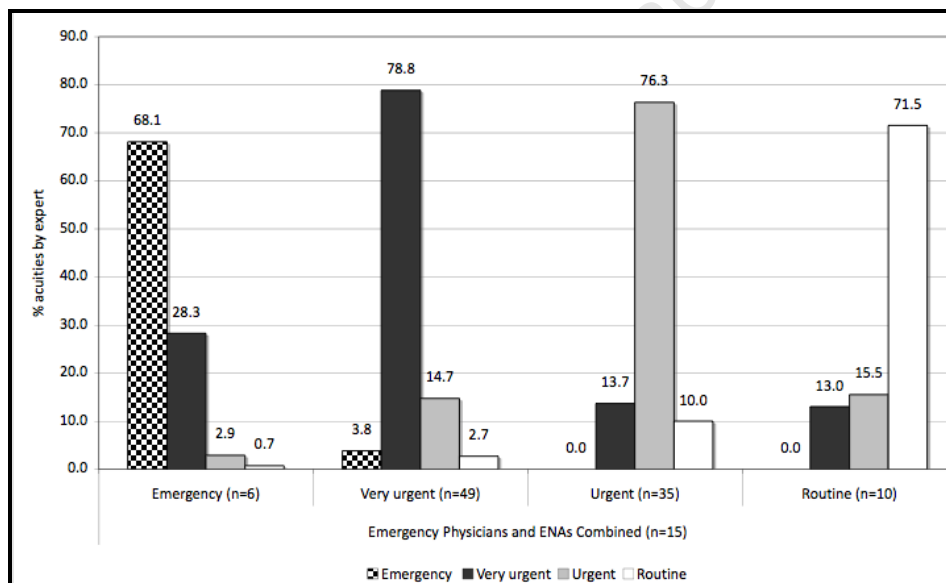


Figure 2: Graphical displays of the PPV, under- and over-prediction of SATS for emergency, very urgent, urgent and routine acuity levels.

Figure 2 summarizes all vignettes according to acuity levels as assigned by the raters, and illustrates the probability that the given ratings are over-/under-predictions of the true acuity (with reference to the experts’ ratings). **Figure 2** indicates, for instance, that about 28% of vignettes triaged “emergency” by the raters were over-predicted by one acuity level, implying that they were actually “very urgent”, about 3% were over-predicted by two acuity levels (i.e. true acuity was “urgent”) and 1% were over-predicted by 3 acuity levels (i.e. true acuity was “routine”).

4. DISCUSSION

This study assessed the validity of the SATS when used by emergency physicians and nurses. The SATS demonstrated good average sensitivity (75%) and specificity (91%). The extent of average over-triage (15%) and under-triage (10%) fell within the given ACSCOT ranges. The average PPV (74%) and NPV (91%) were equally high and the extent of average over-prediction (15%) was higher than under-prediction (11%).

Sensitivity, specificity and percentage over-/under-triage are accuracy summaries (accuracy defined here as the distance from the truth). These summaries may be compared across studies and do not depend on the acuity distribution in a given setting. By contrast, predictive values and over-/under-prediction provide information on actual performance in a given setting. In practice, when a patient is assigned a particular acuity level, the true acuity is not known, and what is of interest from the patient care and resource management perspective is how likely that assigned acuity is to be correct. In our study, the acuities in the set of vignettes represented the distribution at secondary and tertiary level hospital emergency centres.

Under-triage and under-prediction are a concern to patient care, implying longer waiting times, delayed definitive patient care, leading to increased mortality and morbidity. Over-triage and over-prediction do not directly impinge on patient care, but may indirectly compromise patient care for the collective because overstretched and limited resources are diverted from those in genuine need that are truly a higher priority. Over-triage and over-prediction are therefore an important consideration in resource poor settings, where resource allocation, if inappropriately prioritized, may lead to loss of life. Furthermore over-triage/-prediction may result in an over-utilization of finite resources and create an unnecessary financial burden to the health care system, as has been reported in other countries (37). While the extent of under- and over-triage reflects the accuracy of the SATS implementation in a way that is comparable across settings, it is the extent of under- and over-prediction that represents the actual potential impact on patient care and resources in a given setting.

The consequences of mis-assignment of acuity depend on the true acuity level and the extent of mis-assignment. For instance, if the true acuity is “emergency” and the assigned acuity is “very urgent” (an under-triage of the acuity level “emergency” and an under-prediction for the acuity level “very urgent”), then a patient only waits 10 minutes longer for care, whereas if the true and assigned acuity is “urgent” (“routine”) respectively, the patient that should only wait one hour waits three hours longer than appropriate.

4.1 Consequences of mis-triage versus mis-prediction

With reference to the expert triage ratings, we observed a high percentage of under-triage for true “emergency” vignettes (22%). However this was only under-triage by one acuity level implying that

the patient would wait 10 minutes longer, as opposed to an hour or four hours longer. Within the true “very urgent” vignettes 14% were under-triaged, 10% by one acuity level (i.e. 1 hour longer waiting time) and 4% by two acuity levels (i.e. 4 hours longer waiting time). This raised some concern, as the increased waiting times imply compromised patient care. Among the true “urgent” vignettes, 6% were under-triaged by one acuity level (i.e. 3 hours additional waiting time).

Of concern in secondary and tertiary emergency centres is that, of the vignettes triaged as “routine” by the raters, 15.5% were under-predictions by one acuity level (i.e. 3 hours longer waiting time than would be required given the true acuity) and 13% were under-predictions by two acuity levels (i.e. 4 hours longer waiting time). In addition, within the group of vignettes triaged by the raters as “urgent”, 14% were under-predictions by one acuity level (i.e. 1 hour additional waiting time). These summaries reflect the accuracy in practice, for settings represented by the patient mix in this study.

With reference to the expert triage ratings, high over-triage was observed within the true “urgent” (21%) and “routine” (32%) patients.

When using the raters’ triage decisions as a reference, over-prediction was observed in 32% of the vignettes triaged as “emergency” and 17% of the vignettes triaged “very urgent”. Over-prediction of the “emergency” acuity was mostly only by one acuity level, i.e. the true acuities were predominantly “very urgent”, an acuity that requires attention within 10 minutes rather than immediately. Of more concern, in terms of utilization of resources, is over-prediction of the “very urgent” acuity.

The vignettes in this study reflect the distribution of acuities in the population attending secondary level hospitals, where 53% were considered emergency or very urgent by the reference standard, requiring admission. This is similar to previous studies at other urban secondary level hospitals that also have 50% of their EC presentations being admitted as emergency or very urgent patients (10).

The predictive values and percentages of over and under-prediction depend upon the acuity distribution. In general, the positive predictive value of an acuity will tend to increase as its frequency of occurrence increases. In a primary care setting, where the distribution of acuity is skewed towards urgent and routine, potential over-prediction and resource wastage is a concern. Conversely, as one moves higher up the health care referral chain to secondary, tertiary and highly specialised quaternary levels of care, where the distribution is skewed towards higher acuity levels, the PPV will increase at emergency and very urgent acuity levels, and decrease for urgent and routine acuities, indicating that under-prediction is a concern. This emphasizes the desirability of appropriate access at all times to the correct levels of care in an optimally and rationally organised health care system.

5. LIMITATIONS

A limitation of any study assessing the validity of triage scales is the lack of an appropriate reference standard. We chose to address this problem by combining the use of experts and outcome markers in that two triage experts from the SATG made informed triage decisions based on the use of SATS and outcome markers such as hospital admission, death, length of stay and resource utilization. Even though abstracted from real EC presentations, the use of written vignettes was considered a second limitation, as non-verbal cues and visual information can never be entirely accurate in written vignettes, and may have affected the triage decision for the raters.

The reference ranges for average over- and under-triage provided by the ACSCOT are a limited way of interpreting the standard validity indicators in this triage validity study as the reference ranges were developed for pre-hospital trauma triage, they only pertain to average over/under-triage and not to each acuity level, and there are no reference ranges for predictive values.

6. CONCLUSION

The average sensitivity over all acuity levels was 75% and specificity 91%. Average under-triage (10%) and over-triage (15%) with respect to the true acuity fell within the accepted ACSCOT ranges, and confirms previous studies (10) that have shown similar average under-triage (7.8%) and over-triage (13.6%) with respect to the true acuity on SATS performance used among South African nurses. The results of this study indicate that the SATS has good performance characteristics and is a valid scale, which may be further implemented in similar settings in South Africa.

Mis-prediction characteristics are determined by the setting in which triage takes place (primary, secondary or tertiary health facilities) and identify both poor care and/or resource wastage in these different settings.

Further research into accepted norms for extent of over-/under-triage and over-/under-prediction at different acuity levels in the hospital context for both trauma and medical problems is recommended to ensure that appropriate reference ranges and guidelines are available for less developed countries.

7. FUNDING

None

8. CONFLICT OF INTEREST

The authors declare that there is no conflict of interest

9. ETHICS

This study was granted ethics approval from the Research Ethics Committee, University of Cape Town (REC REF: 063/2005).

10. REFERENCES

- (1) Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. *Emerg Med J* 2007;24:477-479.
- (2) Wallis LA. The Cape Triage Score: update. *Emerg Med J* 2006;23:740.
- (3) Mackway-Jones K. Emergency Triage. Manchester Triage Group: BMJ Publishing Group;1997.
- (4) Beveridge R. CAEP issues: the Canadian triage and acuity scale: a new and critical element in health care reform; Canadian Association of emergency physicians. *Journal of Emergency Medicine*, 1998 5/6;16(3):507-511.
- (5) Gottschalk S, Wood D, De Vries S, Wallis L, Bruijns S, On behalf of the Cape Triage Group . The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006 February 1;23(2):149-153.
- (6) Wildschut A, Mqolozana T. Shortage of Nurses in South Africa: Relative or Absolute? 2008. Case study report forming part of the HSRC study: A multiple source identification and verification of scarce and critical skills in the South African Labour market. Available from URL: <http://www.labour.gov.za/documents/research-documents/shortage-of-nurses-in-south-africa-relative-or-absolute>
- (7) Subedar H. Nursing Profession: Production of nurses and proposed scope of practice, In: The South African Health Review 2005. Health Systems Trust (HST). Pretoria.
- (8) Twomey M, Wallis LA, Thompson ML, Myers JE. The South African Triage Scale (adult version) provides reliable acuity ratings. *Int Emerg Nursing*. In press.
- (9) Bruijns SR, Wallis LA, Burch VC. Effect of introduction of nurse triage on waiting times in a South African emergency department. *Emerg Med J* 2008;25:395-397.
- (10) Bruijns SR, Wallis LA, Burch VC. A prospective evaluation of the Cape triage score in the emergency department of an urban public hospital in South Africa. *Emerg Med J* 2008;25:398-402.
- (11) Gottschalk S, Warner C, Burch V, Kow K, Melzer A, Wallis, L. Warning scores in Triage – Is there any point? (Unpublished)
- (12) Streiner DL, Norman GR. *Health Measurement Scales*. 4th ed. Oxford University Press; 2008.
- (13) FitzGerald G, Jelinek GA, Scott D, Gertz MF. Emergency Department Triage Revisited. *Emerg Med J* 2010;27:86-92.

- (14) Fernandes CM, Tanabe P, Gilboy N, et al. Five-level triage: a report from the ACEP/ENA five-level triage task force. *J Emerg Nurs* 2005;31:39–50.
- (15) Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The Emergency Severity Index Triage Algorithm Version 2 is Reliable and Valid. *Acad Emerg Med* 2003;10:1070-1080.
- (16) Taboulet P, Moreira V, Haas L, Porcher R, Braganca A, Fontaine JP, Poncet MC. Triage with the French Emergency Nurses Classification in Hospital Scale: reliability and validity. *Europ J Emerg Med* 2009;16:61-67.
- (17) Wuertz RC, Milne LW, Eitel DR, et al. Reliability and validity of a new five-level triage instrument. *Acad Emerg Med* 2000;7:236-42.
- (18) Wuertz RC, Travers D, Gilboy N, et al. Implementation and Refinement of the Emergency Severity Index. *Acad Emerg Med* 2001;8:170-176.
- (19) Van Gerven R, Delooz H, Sermeus W. Systematic triage in the emergency department using the Australian National Triage Scale: a pilot project. *Europ J Emerg Med* 2001;8:3-7.
- (20) Elshove-Bolk J, Mencl F, van Rijswijck BTF, et al. Validation of the Emergency Severity Index (ESI) in self-referred patients in a European emergency department. *Emerg Med J* 2007;24:170-174.
- (21) Tanabe P, Gimbel R, Yarnold PR, et al. The Emergency Severity Index (version 3) 5-level Triage System Scores Predict ED Resource Consumption. *J Emerg Nurs* 2004;30:22-29.
- (22) Olofsson P, Gellerstedt M, Carlstroem ED. Manchester Triage in Sweden – Interrater reliability and accuracy. *Int Emerg Nur* 2009;17:143-148.
- (23) van der Wulp, Baar van ME, Schrijvers AJP. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. *Emerg Med J* 2008;25:431-434.
- (24) Considine J, LeVasseur SA, Villanueva E. The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Ann Emerg Med* 2004;44:516-523.
- (25) Goransson K, Ehrenberg A, Marklund B, et al. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19:432-438.

- (26) Rutschmann OT, Kossovsky M, Geissb hler A, et al. Interactive triage simulator revealed important variability in both process and outcome of emergency triage. *J Clin Epidemiol* 2006;59:615-621.
- (27) Wollaston A, Fahey P, McKay M, et al. Reliability and validity of the Toowoomba adult trauma triage tool: a Queensland, Australia study. *Acc Emerg Nurs* 2004;12:230-237.
- (28) van der Wulp I, van Stel F. Adjusted weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol* 2009;62:1196-1201.
- (29) Worster A, Sardo A, Eva K, Fernandes CB, Upadhye S. Triage Tool Inter-rater Reliability: A Comparison of Live Versus Paper Case Scenarios. *Journal of Emergency Nursing* 2007;8:319-323.
- (30) Worster A, Fernandes CB. Paper Case Scenarios in the Assessment of Triage Tools. *Annals of Emergency Medicine* 2005;46(2):203-204.
- (31) Worster A, Gilboy N, Fernandes CM, Eitel D, Eva Rose Geisler K, Tanabe P. Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *CJEM: The Journal of the Canadian Association of Emergency Physicians* 2004;6(4):240-245.
- (32) American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient. 1999;98.
- (33) AGREE for Windows [computer program]. Version 7.002 1999. University of Groningen. Available from http://www.scienceplus.nl/component/page,shop.product_details/flypage.shop.flypage_splus/product_id,59/category_id,13/manufacture_id,0/option,com_virtuemart/Itemid,26/ URL:
- (34) Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996;7:199-202.
- (35) Nelson JC, Pepe MS. Statistical description of interrater variability in ordinal ratings. *Statistical Methods in Medical Research* 2000;9:475-96.
- (36) Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *American J Epidemiology* 1987;126:161-169.
- (37) Carr J. The cost of over-triage on our nation's health system. 10/09/2010 13:16:00. Available from URL <http://www.healthcanal.com/public-health-safety/10597-The-cost-over-triage-our-nations-health->

system.html?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+healthnewshc%2FOxfp+%28Health+News+from+HealthCanal.com%29

ADULT TRIAGE SCORE								© South African Triage Group 2008
	3	2	1	0	1	2	3	
Mobility				Walking	With Help	Stretcher/ Immobile		Mobility
RR		less than 9		9-14	15-20	21-29	more than 29	RR
HR		less than 41	41-50	51-100	101-110	111-129	more than 129	HR
SBP	less than 71	71-80	81-100	101-199		more than 199		SBP
Temp		Cold OR Under 35		35-38.4		Hot OR Over 38.4		Temp
AVPU		Confused		Alert	Reacts to Voice	Reacts to Pain	Unresponsive	AVPU
Trauma				No	Yes			Trauma
over 12 years / taller than 150cm								
Colour	RED		ORANGE		YELLOW		GREEN	BLUE
TEWS	7 or more		5-6		3-4		0-2	DEAD
Target time to treat	Immediate		less than 10 mins		less than 60 mins		less than 240 mins	DEAD
Mechanism of injury			High energy transfer					
Presentation			Shortness of breath - acute				ALL OTHER PATIENTS	
			Coughing blood					
			Chest pain					
			Haemorrhage - uncontrolled					
	Seizure - current		Seizure - post ictal					
			Focal neurology - acute					
			Level of consciousness reduced					
			Psychosis / Aggression					
			Threatened limb					
	Burn - face / inhalation		Dislocation - other joint		Dislocation - finger or toe			
			Fracture - compound		Fracture - closed			
			Burn over 20%		Burn - other			
			Burn - electrical					
	Burn - circumferential							
	Burn - chemical							
			Poisoning / Overdose		Abdominal pain			
	Hypoglycaemia - glucose less than 3		Diabetic - glucose over 11 & ketonuria		Diabetic - glucose over 17 (no ketonuria)			
			Vomiting - fresh blood		Vomiting - persistent			
			Pregnancy & abdominal trauma or pain		Pregnancy & trauma			
			Pregnancy & PV bleed					
Pain			Severe		Moderate		Mild	
Senior Healthcare Professional's Discretion								

Appendix 1: The South African Triage Scale

Vignette Description

1: A fragile looking 29 year old woman is assisted into your ED on crutches. She is breathing very rapidly and says that she has been coughing and vomiting for the past few days. She states that she is not in much pain (2/10). She explains to you that she has not been well enough to look after her child. RR 30, HR 160, BP 109/65, 37.5°C

- ☐ Red
☐ Orange
☐ Yellow
☐ Green
☐ Blue

2: "I have had chest pain for 2 weeks now and have a sensation of pins and needles in my left arm and face" reports a 26 year old woman that is carried into your ED on a stretcher. She does not complain of any pain but looks very fearful. RR 23, HR 72 , BP 115/60, 36.3°C, HGT 4.4, HB 11.5

- ☐ Red
☐ Orange
☐ Yellow
☐ Green
☐ Blue

3: Paramedics walk in a 48 year old woman who complains of a moderate headache (5/10) and weakness in her left arm. She is mobile and states that she has had a previous CVA. She is tearful and anxious. RR 20, HR 117, BP 205/136, 36°C, HGT 7.9, HB 10

- ☐ Red
☐ Orange
☐ Yellow
☐ Green
☐ Blue

4: A 26 year old man is brought into your ED by paramedics. He is on a stretcher. He is screaming and shouting about his severe abdominal pain (9/10). He starts throwing his clothes around the ED and becomes disruptive. RR 20, HR 65, BP 131/95, 36.5°C HGT 5.9 HB 14.5

- ☐ Red
☐ Orange
☐ Yellow
☐ Green
☐ Blue

Appendix 2: Acuity assessment – examples of vignettes

Validating the construct of acuity against a set of reference vignettes developed via modified Delphi method

Paper overview

This paper validates the construct of acuity as measured by the South African Triage Scale (SATS), and develops a set of vignettes that may be used as an alternative reference standard for triage scale validation and comparison studies.

Contribution to the thesis and novelty

This paper addresses the sixth and seventh objective of this study by using Delphi methodology to develop a set of reference vignettes. There is currently no other study that attempts to develop a method that will overcome some of the barriers to validation studies in less developed countries. Unique contributions of the paper are the validation of the construct of acuity as measured by the SATS and the presentation of a set of reference vignettes obtained via Delphi method using an international expert panel.

Contribution of candidate

The candidate wrote and managed all drafts of the manuscript, incorporating suggestions from co-authors.

Publication status

Submitted 26 June 2011 for review to Emergency Medicine Journal.

Title:

Validating the construct of acuity against a set of reference vignettes developed via modified Delphi method

Authors:

M.Twomey¹, LA. Wallis², JE. Myers¹

Institutions:

- 1) School of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
- 2) Division of Emergency Medicine, Department of Surgery, University of Cape Town & Stellenbosch University, Cape Town, South Africa

Word Count

Abstract: 255

Paper: 2742

Abstract

Objectives: To validate the construct of triage acuity as measured by the South African Triage Scale (SATS) against a set of reference vignettes.

Methods: A modified Delphi method was used to develop a set of reference vignettes. Delphi participants completed a 2-round consensus building process, and independently assigned triage acuity ratings to 100 written vignettes unaware of the ratings given by others. Triage acuity ratings were summarized for all vignettes and only those that reached 80% consensus during round two were included in the reference set. Triage ratings for the reference vignettes given by two independent experts using the SATS were compared to the ratings given by the international Delphi panel. Standard validity measures including sensitivity, specificity, associated percentages for over-/under-triage were used to estimate the validity of the construct of triage acuity (as measured by the SATS) by examining the association between the ratings by the two experts and the international panel.

Results: On completion of the Delphi process 42 of the 100 vignettes reached 80% consensus on their acuity rating and made up the reference set. On average over all acuity levels, sensitivity was 74% (CI: 64% – 82%), specificity 92% (CI: 87% - 94%), under-triage occurred 14% (CI: 8% – 23%) and over-triage 12% (CI: 8% - 23%) of the time.

Conclusion: The results of this study provide an alternative method for validating the construct of triage acuity as measured by a particular triage scale. This method of using 80% consensus vignettes may however systematically bias the validity estimate towards better performance.

Introduction

With increased overcrowding in Emergency Centres (ECs), triage has been highlighted as a crucial process in prioritizing patients based on medical urgency (1-3). Ideally a triage scale should be highly reliable and valid (4). Reliability tells us how much agreement there is among staff triaging the same patient, while validity tells us how close the assigned acuity rating is to the true acuity of that patient (5).

Much has been published on the validity of different triage instruments (6-16). However, there is a lack of consistency in the statistical measures and reference standards used to report on validity, which makes it difficult to draw comparisons between different triage scales. Before addressing the type of statistical measure used, a reference standard should be introduced when assessing whether the triage tool correctly identifies the true acuity of the patient. Previous discussions in the literature note the fact that triage scales have no unique set of reference values (17-18). Therefore one of the challenges in estimating validity lies in the task of identifying an appropriate reference for the true acuity of the patient (17). Some studies have assessed the validity of triage scales by using either one or two local experts, or outcome markers such as mortality rates, hospital admission rates, resource utilisation and length of stay in hospital (11)(19-20). Although these markers may be adequate in developed countries, they have limitations in less developed countries, which are more adversely affected by factors such as poor record keeping and limited resources. In less developed countries such as South Africa the modified Delphi method has been recommended as an alternative that may be used to develop an objective reference standard for validation studies and overcomes factors such as poor record keeping or ineffective care (17).

Since 1950 the Delphi method has diversified, and has been used more frequently in health care research over the past two decades (21). This consensus building technique is designed to gain insight into a particular field by constructing consensual criteria to enable decision making in areas where published information is scarce or non-existent (22). The approach establishes an appropriate panel of experts that have agreed to partake in an iterative process on a particular issue with the key objective being to reach consensus (23). The Delphi consensus method has been previously used to identify clinical criteria that define triage priority in a major incident setting, and has been recommended for the development of validation criteria for other triage algorithms (24). The South African Triage Scale (SATS) is a 4-level system that objectively categorizes the medical urgency of emergency centre patients based on age-appropriate physiology and clinical discriminators (25). The aim of this study is to validate the construct of triage acuity as measured by the SATS against a set of reference vignettes.

Methods

Study design

This study was conducted in two stages. In the first stage we undertook a modified Delphi method using a two-round consensual process, based on a series of clinical vignettes that had been collected from real EC patient presentations. The source and method of collection of these vignettes have been presented elsewhere (25)(26). In the second stage, we studied the agreement between the two experts' SATS ratings against the reference vignettes as a validation measure for the construct of triage acuity.

Stage One: modified Delphi method

Delphi group

The authors identified international experts in the field of emergency centre triage. 34 experts from developed and less developed countries were approached to take part in a modified two-round Delphi study: 18 (53%) agreed. Participants were selected from countries where triage scales had already been developed and validated or were in the process of being developed and validated. 17 of the 18 participants who agreed to take part came from developed countries. The identity of participants (including emergency physicians and nurses) was kept from the other panel members.

Methods of measurement

Participants were emailed and asked to independently triage written sets of vignettes using the tool they were most familiar with in their day to day practice. The categories they assigned had to fall into one of four triage acuity levels described in **table 1**. Participants were unaware that these were SATS categories. The vignettes were made available online for easy access at a time that was convenient for participants.

Category	Response	Description of category
RED	Immediate	Emergency
ORANGE	Within 10 min	Very Urgent
YELLOW	Within 60 min	Urgent
GREEN	Within 240 min	Routine

Table 1: South African Triage Scale Categories (27)

Data collection

Participants completed a two-round consensus building process, and independently assigned triage acuity ratings to 100 written vignettes unaware of the ratings given by others. Based on extensive use in other studies and its advantage of saving cost and time, vignettes were used as a suitable estimate of live triage cases (28-30). The 100 adult vignettes were prospectively abstracted from randomly selected actual EC case presentations at a secondary hospital in South Africa. Vignettes included

gender, age, presenting complaint, mode of arrival and vital signs. **Appendix A** shows examples. Consensus was built on the acuity ratings given to each vignette and those reaching 80% or more group consensus were included (31). The 80% consensus level is commonly used in Delphi studies and was therefore set as the group consensus level required for inclusion of vignettes into the reference set (31) (32). Participant anonymity was maintained throughout the process, and controlled feedback was provided in the form of a statistical aggregation of the group response after each iterative round (28).

Round 1: Using an online system, Delphi participants were asked to assign acuity categories as indicated in **table 1** to a set of 100 vignettes (see **Appendix A** showing an example of round 1). Any vignettes reaching 80% consensus after round 1 were removed.

Round 2: Vignettes that did not reach consensus were sent back to the Delphi participants. For each vignette the acuity level assigned by the majority was indicated, and participants were given a chance to either change their original acuity rating assigned to each vignette or leave the rating unchanged (**Appendix A** shows an example of round 2). On completion of round 2, triage ratings were summarized for all vignettes and only those that reached 80% group consensus on their acuity rating on either round were included in the set of reference vignettes (31).

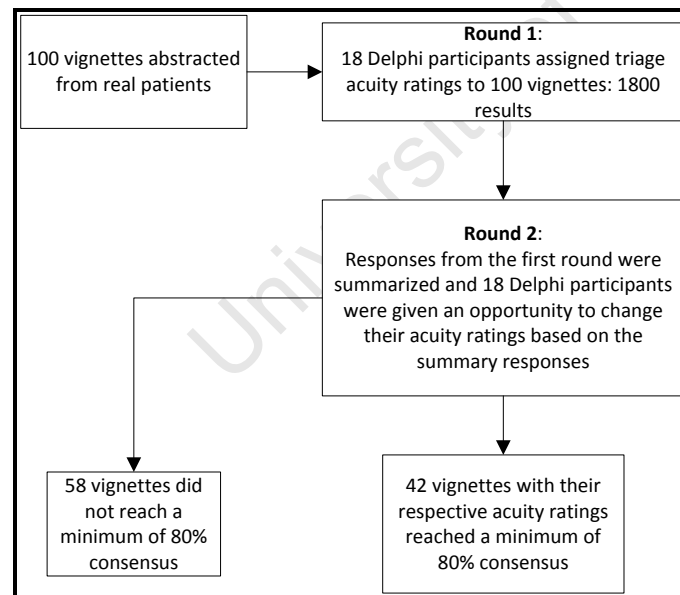


Figure 1: Flowchart of Delphi consensus method

Data analysis

Participant responses of the first and second round of the Delphi process were summarized using STATA statistical software package version 9.2 (33). Descriptive statistics of assigned acuity levels for

the 100 vignettes were summarized for round 1 and 2. Based on previous studies and recommendations in the Delphi literature, only the vignettes that reached 80% consensus were included in the set of reference vignettes (31).

Stage two: Cross-sectional validation of the construct of triage acuity

South African raters

Two South African experts (not part of the Delphi panel) with knowledge and experience of the SATS took part in this cross-sectional validation.

Data collection and analysis

Participants were asked to independently assign acuity ratings to the 42 reference vignettes (as derived in stage 1) using the SATS as their method of triage (**table 1**). The SATS categories assigned by the two South African experts were compared to the reference acuity levels assigned by the Delphi group. Validity was assessed by calculating the sensitivity, specificity, and associated over-/under-triage relative to the Delphi acuity assignments. Histograms were designed to illustrate and visually compare mis-triage at each acuity level.

Results

Modified Delphi

On completion of round two, 80% group consensus was reached on the acuity level of 42 of the 100 vignettes (emergency n=9; very urgent n=17; urgent n=10; routine n=6). **Appendix B** includes the 42 vignettes with their respective acuity ratings that make up the reference set. **Appendix C** includes vignettes that did not reach 80% group consensus, but revealed only one acuity level discrepancy among raters (n=17). The highest group consensus among these vignettes ranged from 61% - 78%.

The vignettes with the lowest plotted percentage group consensus on acuity rating (n=10) are described in **Appendix D**; these ratings were divided across three to four acuity levels. In all ten vignettes the highest group consensus ranged from 50% - 56%. The lack of consensus seen in these ten vignettes pertains mostly to the acuity level 'urgent' indicating systematic bias in the middle acuity categories, which is also visible in **Figure 2**. If we had lowered the consensus cut-off to 78%, there would have been an increase in the vignettes with acuity ratings 'very urgent' and 'urgent', bringing the total number of vignettes with group consensus up to 55. Lowering the consensus cut-off further to 72% would produce an even further increase in vignettes with acuity ratings 'very urgent' and 'urgent', bringing the total number of vignettes with group consensus up to 69.

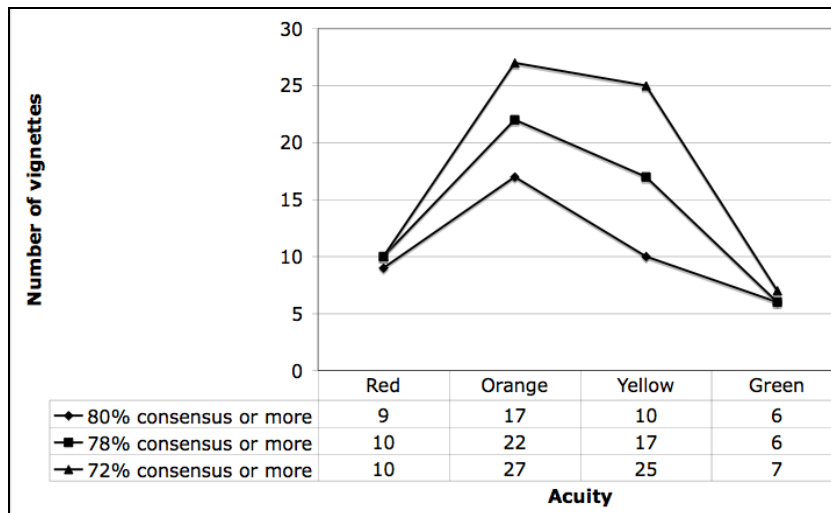


Figure 2: Line graph illustrating different levels of consensus within the Delphi panel

Validation of the construct of acuity

The 42 vignettes that reached 80% group consensus were used as a reference standard to assess the construct of triage acuity as measured by SATS. Two South African experts triaged the 42 vignettes (84 ratings). **Table 2** summarizes the sensitivity analysis and shows that average sensitivity and specificity across all categories was 74% (CI: 64% – 82%) and 92% (CI: 87% - 94%) respectively. Average under-triage across all categories (14%; CI: 8% - 23%), occurred as frequently as average over-triage across all categories (12%; CI: 6% - 20%) (relative to the acuity assigned by the Delphi panel).

Figure 3 summarizes all vignettes with acuity levels as assigned by the South African experts, and illustrates the probability that they will over-/under-triage vignettes at each acuity level relative to the acuity assigned by the Delphi panel. It shows that about 44% of the vignettes given an “emergency” acuity level by the Delphi panel were under-triaged by one acuity level by the South African experts. About 6% of the vignettes categorized “very urgent” by the Delphi panel were under-triaged as “urgent” by the South African experts, and about 30% of the vignettes assigned “urgent” by the Delphi panel, were over-triaged as “very urgent” by the South African experts.

		Vignettes (n)	Triage ratings (n)	2 SA experts combined for 42 vignettes (a total of 84 ratings)				Performance relative to Delphi experts as the objective reference standard			
				Emergency	Very urgent	Urgent	Routine	Sensitivity	Specificity	Over- triage	Under- triage
Delphi panel triage	Emergency	9	18	44.4 *	44.4 z	11.1	0.0	44.4 (CI: 24 – 66)	86.8 (CI: 77 – 93)	n/a	55.6 (CI: 34 – 76)
	Very urgent	17	34	0.0 z	94.1 *	5.9 z	0.0	94.1 (CI: 83 – 99)	94.7 (CI: 85 – 98)	0.0	5.9 (CI: 1 – 17)
	Urgent	10	20	0.0	30.0 z	70.0 *	0.0 z	70.0 (CI: 49 – 86)	90.3 (CI: 82 – 96)	30.0 (CI: 14 – 51)	0.0
	Routine	6	12	0.0	0.0	33.3 z	66.7 *	66.7 (CI: 40 – 87)	94.7 (CI: 88 – 98)	33.3 (CI: 13 – 60)	n/a
Mean								73.8 (CI: 64 – 82)	91.7 (CI: 87 – 94)	11.9 (CI: 6 – 20)	14.3 (CI: 8 – 23)

n/a not applicable

* These values correspond to exact agreement with Delphi expert categories for vignettes

z These values correspond to agreement within one level of the Delphi expert category

Table 2: Summary of SA expert's agreement with the Delphi expert's rating of vignettes, % (confidence intervals)

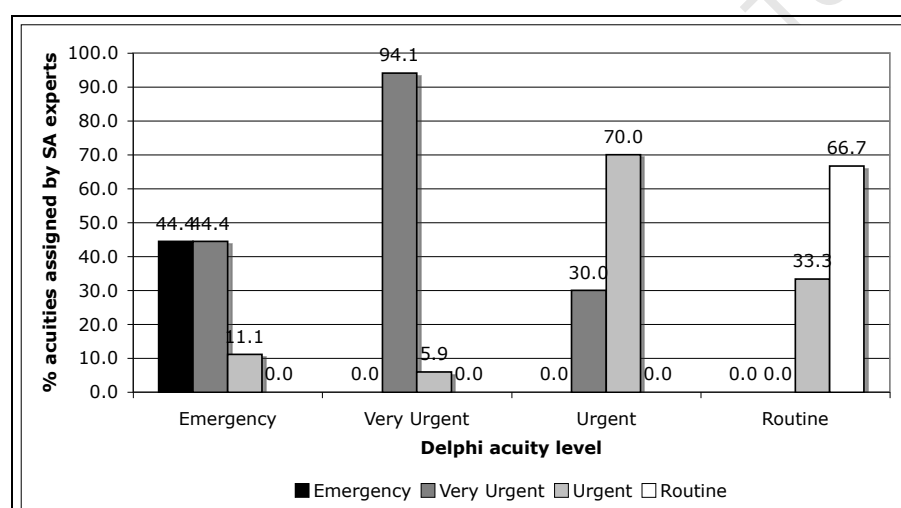


Figure 3: Graphical displays of the sensitivity, over-/under-triage of SATS for emergency, very urgent, urgent and routine acuity levels.

Discussion

This study used a modified Delphi method to build consensus on 100 vignettes and assessed the construct of triage acuity as measured by the SATS. 42 vignettes had a minimum group consensus of 80% and the rest did not meet the minimum group consensus of 80%.

The modified Delphi method may be seen as a form of measuring consensual validation that may be useful when undertaking triage scale validations or comparing triage scale validity in less developed countries. In developed countries triage scales are often validated against outcome markers such as admission, death, length of stay and resource utilisation. These markers assume systematic and

comprehensive record keeping as well as effective care and adequate resources, which in a less developed country may be limited and therefore inappropriate (17).

Validation may be performed using a reference standard of underlying acuity that is consensually arrived at. This has been shown in previous studies where consensus was used from Delphi methodology to establish triage acuity levels against which pre-hospital mass casualty triage tools were tested (32). It has been suggested that emergency centre triage scales be assessed in the same way (17) (24). The benefits of using the results of this modified Delphi process for further triage scale validation studies are that it provides a reference standard that is potentially more appropriate and financially viable for less developed countries.

An accepted performance indicator for measuring validity of triage scales is sensitivity, which measures inherent characteristics of the scale with percentage of over-/under-triage with reference to the Delphi panel ratings in the case of this study. To our knowledge no accepted norms exist for over-/under-triage at each acuity level. The accepted range for average under-triage of not more than 5%-10%, which the American College of Surgeons Committee on Trauma (ACSCOT) considers unavoidable and an associated average over-triage rate of 30%-50% applies exclusively to trauma patients and therefore has limited use when assessing a mixed patient population of trauma and non-trauma cases (34). A recent retrospective cohort study indicated that achieving these ACSCOT benchmarks was not feasible in Pennsylvania, and that these guidelines needed modification if they were to be implemented (35).

The SATS demonstrated satisfactory average sensitivity 74% (CI: 64% - 82%) and specificity 92% (CI: 87% - 94%). The extent of average over-triage was 12% (CI: 6% - 20%) and under-triage 14% (CI: 8% - 23%). High percentages of under-triage are a concern to patient care implying longer waiting times, delayed definitive patient care, leading to increased mortality and morbidity. High percentages of over-triage do not directly impinge on patient care, but may indirectly compromise patient care for the collective because overstretched and limited resources are diverted from those in genuine need that are truly a higher priority. Over-triage is therefore an important consideration in resource poor settings, where resource allocation, if inappropriately prioritized, may lead to loss of life.

Limitations

The selection of the Delphi panel was determined by the research question and may have been limited in its representativeness, as only those experts with an interest may have become involved as participants. We tried to address this by inviting experts from a nursing and emergency physician background from developed and less developed countries. However the majority responded from developed countries. Limitations of the Delphi method are mostly a result of poorly conducted studies rather than fundamental problems. Response rates can be low and may decrease in the

second round (29). However non-response and attrition did not occur in this study as each Delphi panellist personally gave assurance of their participation.

Another potential weakness of the Delphi method is that it may overlook important minority issues because it tries to obtain consensus (36). This method of using 80% consensus vignettes may also systematically bias the validity estimate towards better performance. However despite these limitations we feel that the Delphi method is an appropriate alternative to develop a reference with which to test the validity of triage scales in less developed countries.

Based on extensive use in other studies and its advantage of saving cost and time, vignettes were used as a suitable estimate of live triage cases (28-30).

Conclusion

The results of this study provide an appropriate alternative for triage scale validation and demonstrate that the construct of acuity as measured with the SATS is valid. By utilizing this set of reference vignettes developed via modified Delphi consensus method, we may be able to perform more comparative studies on triage scales in less developed countries and overcome some of the common barriers (such as poor record keeping and resource limitations) experienced.

Acknowledgements

The authors would like to acknowledge the involvement of all participants on the Delphi panel as well as the two local experts from South Africa whose contribution helped make this project possible.

Funding

None

Conflict of Interest

The authors declare that there is no conflict of interest

Ethics

This study was granted approval from the research ethics committee, University of Cape Town (REC REF 063/2005).

References

- (1) Miro O, Antonio MT, Jimenez S et al. Decreased health care quality associated with emergency department overcrowding. *Eur J Emerg Med* 1999;6:105-107.
- (2) Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. *Med J Aust* 2006;184:213-216.
- (3) Spirivulis PC, Da Silva JA, Jacobs IG et al. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *Med J Aust* 2006;184:208-212.
- (4) Iseron KV, Moskop JC. Triage in Medicine, Part I: Concept, History, and Types. *Ann Emerg Med* 2007;49:275-281.
- (5) Streiner DL, Norman GR. *Health Measurement Scales*. 4th ed. Oxford University Press; 2008.
- (6) Bruijns SR, Wallis LA, Burch VC. A prospective evaluation of the Cape triage score in the emergency department of an urban public hospital in South Africa. *Emerg Med J* 2008;25:398-402.
- (7) Gottschalk S, Warner C, Burch V et al. Warning scores in Triage – Is there any point? (Unpublished)
- (8) Storm-Versloot M, Ubbink D, Chin a Choi V et al. Observer agreement of the Manchester triage system and the emergency severity index: a simulation study. *Emerg Med J* 2009;26:556-560.
- (9) Tanabe P, Gimbel R, Yarnold PR et al. Reliability and Validity of Scores on the Emergency Severity Index Version 3. *Acad Emerg Med* 2004;11:59-65.
- (10) Parenti N, Ferrara L, Reggiani MLB et al. Reliability and validity of two four-level emergency triage systems. *Europ J Emerg Med* 2009;16:115-120.
- (11) Olofsson P, Gellerstedt M, Carlstroem ED. Manchester Triage in Sweden – Interrater reliability and accuracy. *Int Emerg Nur* 2009;17:143-148.
- (12) van der Wulp, Baar van ME, Schrijvers AJP. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. *Emerg Med J* 2008;25:431-434.
- (13) Considine J, LeVasseur SA, Villanueva E. The Australasian Triage Scale: examining emergency

department nurses' performance using computer and paper scenarios. *Ann Emerg Med* 2004;44:516-523.

(14) Goransson K, Ehrenberg A, Marklund B et al. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19:432-438.

(15) Rutschmann OT, Kossovsky M, Geissbühler A et al. Interactive triage simulator revealed important variability in both process and outcome of emergency triage. *J Clin Epidemiol* 2006;59:615-621.

(16) Wollaston A, Fahey P, McKay M et al. Reliability and validity of the Toowoomba adult trauma triage tool: a Queensland, Australia study. *Acc Emerg Nurs* 2004;12:230-237.

(17) Twomey M, Wallis LA, Myers JE. Limitations in validating emergency department triage scales. *Emerg Med J* 2007; 24: 477-479

(18) FitzGerald G, Jelinek GA, Scott D et al. Emergency Department Triage Revisited. *Emerg Med J* 2010;27:86-92.

(19) Taboulet P, Moreira V, Haas L et al. Triage with the French Emergency Nurses Classification in Hospital Scale: reliability and validity. *Europ J Emerg Med* 2009; 16: 61-67.

(20) Eitel DR, Travers DA, Rosenau AM et al. The Emergency Severity Index Triage Algorithm Version 2 is Reliable and Valid. *Acad Emerg Med* 2003;10: 1070-1080.

(21) Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ* 1995; 311:376–380.

(22) Pill J. The Delphi method: Substance, context, a critique and an annotated bibliography. *Socio Econ Plan Sci* 1971; 5:57–71.

(23) Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs* 2000;32:1008–1015.

(24) Wallis L, Carly S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. *Emerg Med J* 2006;23:291-95.

(25) Twomey M, Wallis LA, Thompson ML et al. The South African Triage Scale (adult version) provides reliable acuity ratings: a graphical approach enhances conventional methods. Unpublished.

- (26) Twomey M, Wallis LA, Thompson ML et al. Assessing the validity and practical application of the South African Triage Scale (adult version) at secondary and tertiary level emergency centres. Unpublished.
- (27) Gottschalk S, Wood D, DeVries S et al, On behalf of the Cape Triage Group . The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006;23:149-153.
- (28) Worster A, Sardo A, Eva K et al. Triage Tool Inter-rater Reliability: A Comparison of Live Versus Paper Case Scenarios. *Journal of Emergency Nursing* 2007;33:319-323.
- (29) Worster A, Fernandes CB. Paper Case Scenarios in the Assessment of Triage Tools. *Annals of Emergency Medicine* 2005;46:203-204.
- (30) Worster A, Gilboy N, Fernandes CM et al. Assessment of inter-observer reliability of two five-level triage and acuity scales: a randomized controlled trial. *CJEM: The Journal of the Canadian Association of Emergency Physicians* 2004;6:240-245.
- (31) Okoli C, Pawlowski S. The Delphi method as a research tool: an example, design considerations and applications. *Information and Management* 2004; 42:15–29.
- (32) Wallis LA, Carley S, Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. *Emerg Med J* 2006; 23:291–295.
- (33) Intercooled Stata statistical software package version 9.2 for Macintosh, 21 Nov 2006. Stata Corporation 4905 Lakeway Dr College Station, TX 77845 USA. Available at <http://www.stata.com/>
- (34) American College of Surgeons Committee on Trauma. Resources for Optimal Care of the Injured Patient. 1999;98.
- (35) Mohan D, Rodengart MR, Farris C, et al. Assessing the feasibility of the American College of Surgeons' Benchmarks for the triage of trauma patients. *Arch Surg* 2011; 43: E1-E7.
- (36) Van Teijlingen E, Pitchforth E, Bishop C et al. Delphi method and nominal group techniques in family planning and reproductive health research. *J Fam Plann Reprod Health Care* 2006;32:249-252.

APPENDIX A: Round 1 & 2 - Acuity assessment: Examples of vignettes and the summary response respectively

ROUND 1: Vignette Description

1: A fragile looking 29 year old woman is assisted into your ED on crutches. She is breathing very rapidly and says that she has been coughing and vomiting for the past few days. She states that she is not in much pain (2/10). She explains to you that she has not been well enough to look after her child. RR 30, HR 160, BP 109/65, 37.5°C

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

ROUND 2: Vignette Description with summary response

1: A fragile looking 29 year old woman is assisted into your ED on crutches. She is breathing very rapidly and says that she has been coughing and vomiting for the past few days. She states that she is not in much pain (2/10). She explains to you that she has not been well enough to look after her child. RR 30, HR 160, BP 109/65, 37.5°C

The majority of the Delphi panel assigned this vignette with an acuity rating “RED / emergency”, you assigned an acuity rating “Orange / very urgent” – please mark your rating?

- ☐ Red
- ☐ Orange
- ☐ Yellow
- ☐ Green
- ☐ Blue

APPENDIX B: Set of 42 vignettes that achieved 80% consensus on acuity rating.

	RED / EMERGENCY VIGNETTES (n=9)
1	Paramedics were called to see an acutely disorientated 70 year old man. He became acutely short of breath at the same time, an hour or so ago. They bring him to you on a stretcher. He now only responds to voice. He was independent at home, but hasn't been seen for 2 days. He is unkempt and his clothes are tattered. RR 44, HR 163, BP 113/69, 36.4°C, HGT 12.4, HB 9
2	An unresponsive 33 year old male is rushed into your ED on a stretcher by the paramedics. His sister says that she found him unconscious in bed this morning. His body appears to be going into episodic spasms. He smells of alcohol, is unkempt and has been incontinent of urine. His sister reports that their brother had unexpectedly passed away 2 weeks ago. RR 26, HR 138, BP 79/60, 37°C, HGT 3.1, O2 SATS 95%
3	Paramedics carry a 33 year old man into your ED on a stretcher. He attempted suicide by jumping out of a fourth floor window. There is dried blood in his face and around both legs. He is immobilised on a spinal board and says his body is in severe pain (8/10). He appears to have multiple fractures of his lower limbs. RR 20, HR 88, BP 105/58, 37.2°C, HGT 10.2
4	A 32 year old woman comes into your ED leaning on her husband's shoulder. She is a known asthmatic and her spouse says that he is worried because she is very short of breath. She is alert but too breathless to talk. She looks pale and sweaty. RR 34, HR 108, BP 151/78, 37.4°C, HGT 5.9, HB 9
5	Paramedics bring a 65 year old man into your ED on a stretcher. He has a gun shot wound to the right upper quadrant of his abdomen. He looks weary, pale and complains of mild pain in the right iliac fossa. There is blood on his shirt over the entry wound. RR 20, HR 84, BP 147/91, 37°C, HGT 5.5, HB 12
6	A 56 year old exhausted woman is brought into your ED with the help of her husband. She is groaning, very weak and unable to walk. Her face is very pale and sweaty. She says she is short of breath and in severe body pain (8/10). RR 29, HR 123, BP 84/57, 37.5°C, HGT 8.3, HB <4, O2 SATS 96%
7	A badly injured 46 year old man is rushed into your ED on a stretcher. Paramedics report that he was drinking heavily all day, then was knocked over by a car. He is bleeding from several areas on his body and responds to voice only. He appears to have multiple injuries to his head, arms, shoulders and legs. RR 26, HR 83, BP 143/83, 37°C, HGT 7.2, HB 12.5
8	An immobile 35 year old man is brought into your ED on a stretcher. He has a past history of pulmonary TB. He is a known alcohol abuser and has been confused and unable to speak for the past 2 days. He is unkempt and his clothes are grubby. He responds to pain only. RR 23, HR 148, BP 129/83, 36.8°C
9	A cold and sweaty 33 year old man is helped into your ED by paramedics. He is seated in a wheelchair, responds to voice only and the paramedics state that he was stabbed over his left scapular and in his right leg. He appears to be in severe pain (8/10). The paramedics state that the bleeding of both wounds is now controlled. RR 30, HR 100, BP 116/74, 37°C

	ORANGE / VERY URGENT VIGNETTES (n=17)
10	A maximum security prisoner aged 29 years walks into your ED accompanied by security guards. He says he is short of breath and complains of weight loss. He looks pale, tired, weak and appears to be immuno-compromised with a history of Pulmonary TB. RR 23, HR 72, BP 115/60, 36.3°C, HGT 4.4, HB 11.5
11	"This is the worst pain I've ever felt" exclaims a 42 year old male who walks into your ED with difficulty. He looks very uncomfortable and is pale. On a scale of 1 to 10 he rates his left buttock pain as 9. He has had a previous buttock abscess and also has pulmonary TB. RR 16, HR 73, BP 150/85, 37°C, HGT 5.8, HB 12
12	"I have just been to Mecca and since the flight back a few days ago I have had such severe chest pain (7/10)" reports a 65 year old man. The pain is central and heavy. He is mobile, does not look distressed but is clammy and pale and strangely despondent. He is a known diabetic. RR 36, HR 89, BP 116/83, 37°C, HGT 19.2, HB 10
13	An 18 year old female is brought into your ED in a wheelchair by her mother. Her mother says that she fell to the ground in a shopping mall after fainting. She is now bleeding from her right eardrum and vomiting persistently. She is alert but deeply unhappy. She looks extremely pale, refuses to eat and is 13 weeks pregnant. She tells you she "feels awful" and vomits on the floor of the triage area. RR 17, HR 80, BP 97/61, 37°C, HGT 5.5, HB 9
14	A 76 year old female walks into your ED. She looks well but says that she experienced sharp central chest pain last night. She lives alone and was transported to the ED by her 23 year old grandchild. Now she states her chest pain as being moderate (6/10). RR 16, HR 55, BP 150/68, 36.5°C, HGT 7.2, HB 9
15	A 35 year old woman walks into your ED. She looks exhausted and complains of a cough and wheezing. For the last 10 days she hasn't been sleeping well and has been feeling weak. She is a smoker with no past history of note. RR 30, HR 125, BP 115/67, 38°C, HGT 4.8, HB 11
16	A 78 year old woman is assisted into your ED in a wheelchair. She has been vomiting and complains of severe abdominal pain (9/10). She looks unwell, and is a known diabetic on insulin. She hasn't eaten properly for two days and looks anxious to you. RR 28, HR 95, BP 147/75, 36°C, HGT 9.3, HB 10.5
17	"I had such severe chest pain last night" says a 72 year old woman as she walks into your ED. She has a history of hypertension. She looks well to you and states that her pain is now minor (3/10). RR 12, HR 102, BP 180/94, 36.5°C
18	A frail looking 27 year old HIV+ woman is brought into your ED in a wheelchair. She is very confused, responding to voice only and has a severe headache (8/10). She is accompanied by her brother but looks uncared for. RR 12, HR 77, BP 111/66, 37°C, HGT 5, HB 6
19	A 36 year old man is accompanied into your ED by his concerned wife. He is mobile but appears to be in extreme agony. He says that he has a fever and severe headache (9/10). He lies down on a bed, holding his head in pain. RR 18, HR 95, BP 123/84, 37.8°C
20	Paramedics bring a 55 year old male with moderate chest pain (6/10) into your ED. He has a family history of cerebral-vascular accidents and also complains of ear pain. He is mobile, appears well presented, is softly spoken but does not look well to you. RR 12, HR 96, BP 152/90, 36°C, HGT 4.4, HB 12
21	A 63 year old well-looking female walks into your ED accompanied by her husband. She complains of mild chest pain (3/10) that started 2 hours ago. Her husband expresses tremendous concern but she does not appear to be very distressed. RR 20, HR 74, BP 124/73, 36°C, HGT 6, HB 11
22	A 57 year old man walks into your ED with severe abdominal pain (7/10). He says that he had an alcohol binge 2 days ago and he is vomiting. He becomes tearful and tells you that he is attending an alcoholics anonymous support group but can't stop himself from relapsing after being sober for 10 years. RR 16, HR 111, BP 218/137, 36°C, HGT 7.9, HB 13
23	A 67 year old diabetic man walks into the ED looking fit and healthy. He says that he has moderate chest pain (5/10) that started 6 hours ago. He tells you that his father died a few days ago and that he has been very emotionally drained and tired. RR 24, HR 69, BP 122/69, 36°C, HGT 8.7, HB 11
24	"I was shifting my bed and now have moderate left sided chest pain (5/10)" states a 93 year old very well looking woman who comes into your ED in a wheelchair. She is accompanied by her daughter but lives alone and has no history of other medical problems. She does not appear to be very concerned about her injury and keeps telling you jokes. RR 12, HR 83, BP 171/89, 36.8°C
25	"I injured my back at work and since then have been vomiting and feeling dizzy" states a pale looking 25 year old man. He is mobile but looks very concerned and has no past history of other medical problems. He vomits in the triage area and you notice it contains some fresh blood. RR 20, HR 119, BP 126/89, 36.6°C, HGT 5.3, HB 8
26	A 39 year old woman states that she is experiencing the worst abdominal pain ever (9/10). She is mobile and alert, but appears to be in extreme agony. She is accompanied by her husband who tells you that she has had a previous surgical evacuation of her uterus following the recent birth of their child. RR 12, HR 91, BP 134/91, 37°C, HGT 4.5, HB 9

	YELLOW / URGENT VIGNETTES (n=10)
27	A 40 year old woman walks into your ED complaining of moderate lower abdominal pain (6/10). She has a heavy PV bleed, but feels well with it. She looks comfortable but is very hungry and requests some food. RR 18, HR 85, BP 148/89, 36.5°C
28	A very talkative 57 year old man walks into your ED with extremely swollen legs and a distended abdomen. He has a history of hypertension and COAD but nothing else of note. He complains of moderate pain (5/10) in his legs and was well until his legs swelled up. RR 16, HR 106, BP 126/77, 36.8°C, HGT 11.2, HB 12.5
29	"I have just returned from Egypt and am just not feeling right", reports a 39 year old male. He is mobile and complains of a moderately severe headache (6/10) but can't be more specific about his other symptoms. He is very concerned and fearful as are his family: they all think he has malaria. RR 20, HR 102, BP 125/81, 38.4°C, HGT 4.8, HB 10.5
30	A 35 year old pale and emaciated woman walks slowly and with difficulty into your ED. She is coughing and complains of loss of appetite, weight loss and diarrhea. She is known to be HIV positive, is on ARV treatment and hasn't been well for a couple of weeks now. She desperately tells you how her family and community have disowned her since becoming HIV positive. RR 20, HR 111, BP 100/70, 38.5°C
31	A 62 year old woman with an arm injury walks into your ED accompanied by her husband. She states that she fell over her cat this morning and is now in moderate pain (6/10). It appears that she has fractured her left wrist. She starts chatting to the nurse and asks for a cup of tea to calm her nerves. RR 20, HR 52, BP 121/78, 36.9°C
32	A distressed 34 year old man walks into your ED complaining of moderate abdominal pain (6/10). He is well dressed and orderly but looks very anxious. He is accompanied by his girlfriend who does not stop expressing her concern for his health. RR 12, HR 88, BP 157/97, 37.1°C
33	Paramedics bring a 57 year old woman into your ED stating that she has had four tonic seizures. She is alert on arrival, mobile, has an abrasion on her lip but no focal neurology. She states that this happens frequently and that her family takes such good care of her but that she was walking to the shop without anyone accompanying her. RR 12, HR 67, BP 176/92, 36.2°C, HGT 5.4, HB 12.8
34	A very distressed 74 year old man walks into your ED complaining of discomfort in his groin and hip. He has an enlarged prostate and rates his pain as moderate (6/10). His wife states that he has never complained about discomfort this much before. RR 20, HR 76, BP 164/88, 36°C
35	A 46 year old fit looking woman walks into your ED with severe pain (7/10). She says she fell and injured her arm. She has no history of other medical problems. She describes to you how she fell from her bicycle and had to wait half an hour before anyone stopped to help her. RR 17, HR 82, BP 164/110, 36.4°C, HGT 5, HB 11
36	A thin 21 year old male walks into your ED with an injury to his left eye. He complains that it is quite painful (6/10) and that it feels like something is stuck in it. He claims that he was angle grinding but cannot quite recall what happened. RR 20, HR 79, BP 142/79, 36.2°C

	GREEN / ROUTINE VIGNETTES (n=6)
37	A 22 year old healthy looking woman walks into your ED. She is concerned about her face being swollen and red after spending a day in the sun. She is very chatty and says that her face is also a little itchy. RR 18, HR 80, BP 141/85, 36°C
38	A 20 year old pregnant female walks into your ED for a review. She has a history of abdominal pain during pregnancy but is otherwise healthy and happy. She states that she would like a check up as it has been some time since her last appointment. She has had mild abdominal pain (3/10) in the past month. RR 12, HR 87, BP 126/68, 36.8°C
39	A healthy looking 26 year old male walks into your ED. He says that he has mild right loin pain (4/10). He is well dressed and states that he has no other history of note. He claims to lead a very healthy lifestyle as he is a professional opera singer. RR 12, HR 67, BP 157/98, 36°C
40	A very healthy looking 16 year old female says that her earring is stuck in her ear and she is not able to remove it. She is mobile and has no other past history of note. She is extremely embarrassed when her mother tells you that no one at home was able to help remove the earring so they decided to come to the ED. RR 18, HR 89, BP 116/74, 36.5°C
41	A fit and healthy looking 35 year old man walks into your ED stating that he hit his thumb with a hammer. He does not appear to have a dislocation and is not particularly distressed. He proudly tells you that although he hurt his thumb he was still able to fix his cupboard. RR 20, HR 68, BP 145/92, 36°C, HGT 4.9, HB 12
42	A very healthy and fit looking 17 year old female walks into your ED complaining of hoarseness. She does not appear to be distressed but says that she had to take the day off school to come to the ED because her drama teacher told her that she has vocal nodules. She is afraid that she will need an operation. RR 16, HR 62, BP 121/71, 36.5°C

APPENDIX C: Vignettes that did not reach 80% group consensus revealing one acuity level discrepancy. R= red, O= orange Y= yellow, G= green

%	Description of patient with presenting complaint and vital signs (n=17)
Y= 78 G= 22	A 39 year old female walks into your ED. She is well dressed, looks comfortable, but complains of moderate pain in her right leg (5/10). She does not appear distressed and states that she takes Warfarin and has previously had a DVT. She has no other past history of note. RR 14, HR 76, BP 119/77, 37°C
O= 78 Y= 22	A 40 year old man walks out of an ambulance and into your ED. He is breathing rapidly and looks uneasy. He was working with acid earlier in the day, and has been breathless for “a couple of hours”. He is an asthmatic. RR 19, HR 93, BP 133/72, 37.2°C, HGT 5.3
R= 67 O= 33	Paramedics carry a 59 year old known epileptic male into your ED on a stretcher. He has just had a seizure and responds only to pain. He does not have any apparent injuries and looks well dressed and orderly. RR 25, HR 163, BP 150/81, 36.4°C, HGT 6.6, HB 12
O= 61 Y= 39	A drowsy 40 year old woman is assisted into your ED after taking an overdose of Benzodiazepine and cocaine. She is a known drug addict with a history of depression. She is alert but uncooperative and a little confused. She has dark bags under her eyes and looks scruffy. RR 12, HR 65, BP 120/80, 37°C
Y= 67 G= 33	A 24 year old male is brought into your ED by paramedics with an old left hand injury that has become septic. He is alert, mobile and otherwise very healthy. You see that his hand is swollen and red after removing his bandage. RR 14, HR 85, BP 166/100, 37°C, HGT 4.9
O= 67 Y= 33	A 58 year old man has been fitting for 6 hours on and off pre-arrival. He is brought into your ED on a stretcher. He is alert at triage, but really doesn't feel so well. He is unkempt and does not appear to have any obvious injuries. RR 23, HR 67, BP 126/81, 36.9°C, HGT 5.2, HB 11.6, O2 SATS 100%
G= 72 Y= 28	A frail looking 91 year old woman is transported from her old age home to your ED. She walks with assistance and complains of mild left hip pain and left ankle pain (3/10). She has not experienced any recent trauma but fell 5 years ago and has an old hip injury. She is very chatty but a little confused, which is normal for her apparently. RR 16, HR 65, BP 127/64, 37.4°C, HGT 4.1, HB 9
O= 72 Y= 28	A 24 year old woman walks into your ED after being referred by her GP. She is sweaty and appears to be in a lot of discomfort. She states that she is feeling really unwell, has moderate to severe backache and fever (7/10). She has felt really unwell for 2 days now. RR 16, HR 114, BP 95/60, 40°C
R= 78 O= 22	A weak looking 40 year old woman walks into your ED stating that she has severe chest pain (7/10) and is short of breath. She explains that she has experienced a strange tingling sensation in her arm twice in the past month and sometimes feels a little dizzy too. RR 42, HR 131, BP 127/74, 38°C, HGT 7, HB 6
G= 61 Y= 39	A 26 year old drunk male, wearing torn clothes is brought into your ED by paramedics. The patient complains of a swelling in the left groin area. He is a known alcoholic that lives at a nearby night shelter. He requests to take a shower and asks the nurse for some food. RR 18, HR 83, BP 124/72, 36.3°C, HGT 4.9, HB 10
R= 61 O= 39	Paramedics bring a 42 year old HIV+ woman into your ED. She is sweaty, distressed and immobile. She has difficulty speaking as her breaths are very shallow and rapid. She looks frightened and pale. RR 32, HR 140, BP 108/73, 39.3°C, HGT 4, HB 8.5
Y= 61 G= 39	A 45 year old man with moderate pain (5/10) above his left eye walks into your ED stating that he was assaulted 3 days ago. He has been referred from the day hospital for a review and has not experienced any new trauma since he was hit with a stick on his eyebrow. RR 20, HR 55, BP 146/43, 36°C
O= 72 Y= 28	Paramedics assist a very distressed 51 year old male into your ED. He says he has severe lower back pain (9/10) which he feels all the way down his leg. He is able to stand but cannot walk. This patient has a past history of chronic backache and spine operations. He lies very still and groans in pain. RR 17, HR 72, BP 145/81, 36.2°C
Y= 61 O= 39	A very vocal 32 year old man walks into your ED stating that he was allegedly assaulted. He complains of severe chest pain (8/10) after being kicked in the left rib and has a productive cough. He tells you that his family was being attacked at home and that he is very grateful that his wife and children are still alive. RR 20, HR 106, BP 173/101, 35°C
Y= 78 G= 22	“I had a right knee operation 10 days ago and am now experiencing moderate pain (6/10) and swelling in my right knee” reports a 52 year old woman. She is seated in a wheelchair and tells you that she has been resting her knee for the past few days but the pain seems to be getting worse. She has an out patient appointment in 2 days time. RR 16, HR 112, BP 155/99, 36°C
Y= 72 O= 28	A 66 year old man comes into your ED with a tight chest and wheezing. He appears to have an exacerbation of COPD but is in no pain and does not appear to be very distressed. He is mobile and accompanied by his wife who claims that she is very concerned about the fact that he does not stop smoking. RR 24, HR 84, BP 112/81, 37°C, HGT 7.9, HB 12
R= 61 O= 39	A fragile looking 29 year old woman is assisted into your ED on crutches. She is breathing very rapidly and says that she has been coughing and vomiting for the past few days. She states that she is not in much pain. She explains to you that she has not been well enough to look after her child. RR 30, HR 160, BP 109/65, 37.5°C

APPENDIX D: Vignettes with the lowest % group consensus (n=10) R= red, O= orange, Y= yellow, G= green

Description of patient with presenting complaint and vital signs (n=10)	% acuity distribution			
	R	O	Y	G
A 58 year old man limps into your EC with a left toe infection. He is in no pain but has been vomiting persistently. He looks very tired, does not say much when asked about the length of his infection and appears very restless. His clothes are torn and grubby. RR: 16, HR: 166, BP: 118/78, 37°C, HGT: 17, HB: 10	50	28	22	0
A 48 year old man walks into your EC complaining of moderate chest pain and headaches (5/10). He reports that he has been experiencing a great deal of stress at work and that he lost his wife to cancer 2 weeks ago. RR: 20, HR: 97, BP: 109/73, 36.6°C, HGT: 4.7, HB: 13	0	55.5	39	5.5
Paramedics carry a 20 year old male with multiple abrasions on his knees and shoulders into your EC. He is lying on a stretcher and reports that he was hit by a car whilst cycling. The injured man states that he is in moderate pain (6/10). RR: 12, HR: 71, BP: 130/61, 37.2°C	5.5	28	55.5	11
A 28 year old woman walks into your EC complaining of a painful swelling on her right breast. She rates her breast pain, which appears to be an abscess as 9 on a scale of 10 and states that she also has herpes zoster. She has previously had a right breast abscess and you suspect that she is HIV positive. RR: 13, HR: 86, BP: 122/87, 38°C, HGT: 4.2, HB: 7.5	0	56	33	11
"My boyfriend had too much alcohol last night and hit me over the head" reports a 19 year old female as she walks into your EC. There is some dry blood around the laceration on her head but it appears to be superficial. You notice some cigarette burns on her arms and legs. RR: 12, HR: 88, BP: 106/68, 36.4°C, HGT: 5.3, HB: 9	0	11	50	39
A 27 year old man comes into your EC in a wheelchair. His foot is covered in blood and he states that it got caught in a machine at work. He complains of moderate pain (6/10). He has no past history of note and looks healthy. RR: 18, HR: 49, BP: 135/69, 36°C	0	33	56	11
A 30 year old exhausted woman is brought into your EC with the help of paramedics. She complains of moderate abdominal pain and vaginal discharge. The pregnancy test that is done comes out negative. She looks disappointed and tells you that she has been struggling to fall pregnant. RR: 21, HR: 72, BP: 121/73, 38.5°C, HGT: 6.8, HB: 10.5	0	6	50	44
A 30 year old HIV+ man walks into your EC complaining of a cough and fever. He is a smoker with a swollen right lymph node and says that he is not in much pain (2/10). He looks comfortable and relaxed. RR: 36, HR: 137, BP: 104/63, 39.8°C	5.5	33	56	5.5
A 30 year old fit looking man with known renal artery stenosis walks into your EC complaining of intermittent hand and foot swelling. He says he is in no pain but claims that the hand swelling affects his ability to work and annoys him. RR: 16, HR: 94, BP: 149/83, 37.2°C	0	6	44	50
A very frail and dehydrated 79 year old woman is brought into your EC by paramedics. She is seated in a wheelchair and complains of persistent vomiting and mild abdominal pain (5/10). She is a known diabetic and has a past history of renal failure. RR: 18, HR: 76, BP: 209/76, 37.4°C, HGT: 7.8, HB: 8	5.5	39	55.5	0

Chapter 3: Discussion

This thesis has resulted in a series of five linked papers which all contribute to addressing the original aims of the thesis. The synopsis below reflects on how the papers have contributed to addressing these. In order not to repeat the discussion points raised in each article the synopsis attempts to draw out themes from the integrated body of work, and updates discussion points where additional data have become available subsequent to the publication of the papers included in the thesis. The synopsis also draws on data from related work by the student and colleagues.

Reliability

Despite a large amount of triage literature in the public arena and the recent proposal of guidelines for reporting reliability and agreement studies,⁸⁴ there is still neither consensus nor clarity as to the most appropriate measure for assessing triage reliability. This lack of clarity is largely attributed to the complexity of the exercise itself within varying contexts. Each context will differ depending on resource availability (i.e. number and level of skilled human resources, structure, equipment and stock), number and acuity of patients. Thus EC triage is dynamic and will function on a continuum according to contextual needs. It requires an appropriate assessment of its performance characteristics in the respective context. The interchangeable use of conceptually distinct terms in the literature, (such as reliability and agreement) create methodological uncertainty. The paucity of published emergency medicine triage literature on the conceptual distinction between reliability and validity leads to further confusion when assessing performance characteristics of a triage scale.

Limitations of the quadratically weighted kappa and how to overcome them

While the quadratically weighted kappa statistic is the most commonly reported statistic in the triage assessment literature, the literature reports that the kappa statistic suffers from the disadvantages of its dependence on the number of categories in the ordinal data⁵⁸, the frequency distribution of the mix of cases by severity, and its being a general estimate of agreement over all categories of severity, thereby losing information about agreement for each specific category.^{49 59} This limits generalisability across study settings.^{49 58} Summary measures of reliability, particularly the kappa statistic, inadequately adjust for differential study characteristics, making meaningful comparison across reliability studies problematic.⁶⁰

This thesis and findings from a further reliability study from Botswana⁸⁵ describe a graphical method for examining inter-rater reliability for EC triage tools. This provides useful additional information relating to specific triage urgency categories over and above the single measure of agreement over all urgency categories provided as a summary measure by a kappa statistic. Triage tool reliability assessment may

benefit from this type of visual presentation, especially when comparing agreement around specific urgency categories between study settings. The findings of this thesis provide acceptable reliability of the SATS, whether it is used by ENAs, final year medical students or emergency physicians, and provide supportive evidence for further implementation of the SATS in similar settings by health professionals with different levels of training.

Limitations of the software packages

Quadratically weighted kappa measures for multiple raters cannot currently be calculated using STATA, SPSS or SAS. Based on reports and recommendations in the literature, AGREE 7 for Windows⁸⁶ appeared to be an alternative that provided an option to calculate the quadratically weighted kappa for multiple raters.⁸⁷ However the package was difficult to use, and was characterised by limited documentation and user support.

Recently a novel adjusted weighted kappa for severity of mis-triage was proposed as an alternative, more appropriate, measure for triage reliability.⁸⁷ The weighting scheme in this measure was designed to take into consideration the severity of over-triage and under-triage so as to reflect clinical triage practice. The authors introduce differential weightings for over- versus under-triage based on the quadratically-weighted kappa.⁸⁷ Over- and under-triage can, however, only be defined in relation to an objective reference and hence this adjusted weighted kappa was not considered to be a reliability measure. The adjusted weighted kappa for severity of mis-triage is a measure of *validity* that uses different standards, thereby limiting its utility for comparisons between validation studies. It cannot be used for assessing reliability. This highlights that the interchangeable use of two conceptually distinct terms (i.e. reliability and consistency) may obscure the underlying methodology and further blur conceptually distinct characteristics such as reliability and validity.

Validation

Current methods of triage scale assessment and their limitations

Predictive validity (a type of construct validity) is the most frequently used method of developing, assessing and further refining triage tools.²² It considers the degree to which the triage acuity level is able to predict true acuity. Particular outcomes, or events with time-ordering, are selected as surrogate markers (such as mortality rates, hospital admission rates, resource utilization and length of stay in hospital). There are methodological problems with the use of this type of validity as it does not always answer the core question “Is the triage instrument able to measure what is supposed to be measured?” In patients it does

not measure acuity at the time of assessment (and is inherently confounded by the effectiveness of the health care intervention).

Examples of predictive validity are commonly found in the triage literature, as surrogate outcome markers are practical to measure and are claimed to be closely associated with true acuity.^{22-23 42} This has compelled clinicians and researchers to utilise triage instruments as prediction tools. However our ability to identify and measure the relationship between patient acuity level and outcome depends not only on the measurement of the surrogate outcome marker and the patient's acuity level, but also very importantly on confounding variables such as variability in triage nurse decisions, and delayed and ineffective treatment. These may affect the surrogate outcome marker.

How can limitations be overcome?

A detailed literature review revealed that very little has been published on triage validity in developing countries. The World Health Organization (WHO) reports that triage research is not a priority in low- to middle-income countries.⁸⁸ They have accordingly developed the Emergency Triage Assessment and Treatment (ETAT)⁸⁹ for application to developing countries. While this subjective system has been successfully implemented in Malawi, countries like India, Brazil and South Africa have sought a more objective triage instrument based on physiology. They have either adopted the triage instrument from a developed country or modified it to their own local context and needs.⁹⁰ South Africa has adapted the Modified Early Warning Score (MEWS) to the Triage Early Warning Score (TEWS) and developed the South African Triage Scale after validating it on the local population.¹⁰ Some areas of Brazil have adopted the CTAS, others the ESI.⁹⁰

A triage tool designed for a developed country may be valid in that context leading to favourable results that are meaningful and have implications for action. If, however, the same triage tool were applied in a developing country results may vary due to different resources and skills. Similarly results may vary when applying surrogate markers from developed countries to undertake validity testing in developing countries. This variability may increase the random error in both triage acuity level and outcome category, it would therefore be more appropriate to apply a locally developed tool that is meaningful in the local context.

Whichever tool is used, an assessment of its usefulness in these settings is required. When selecting surrogate outcome markers (such as mortality rates, hospital admission rates, resource utilization, and length of stay in hospital), it is assumed that there is systematic record keeping, and that the care given is effective. While this may often be the case in developed countries, it is typically not the case in less developed countries. Poor record keeping and ineffective care may have significant effects on surrogate outcome markers and patients' final dispositions. Markers such as these are imperfect measures of patient acuity in the developing world. It is thus important to identify and measure all confounding variables that

may be affecting the surrogate outcome marker: given the poor record keeping and lack of efficiency, this is unlikely to be feasible in developing countries. Other possible indicators that have not been extensively explored are patient pathology and/or a list of accepted interventions for each respective acuity level. These indicators would need to be developed via consensus method of which the Delphi method is an accepted technique.

Delphi Methodology

The Delphi technique has diversified and is being applied within the healthcare arena.⁹¹ It is a consensus building method that may be useful when assessing triage scales in less developed countries. It allows the development of a reference standard determined by specialist panel consensus. The triage tool's validity may then be tested against this to validate the construct of acuity as measured by a particular triage scale.

Wallis et al.⁹² used consensus from Delphi methodology to establish triage acuity levels against which to test pre-hospital mass casualty triage tools: such methodology may be used in EC triage tool assessment. There are several reasons why the Delphi methodology is best suited to assessing EC triage tools in developing countries. The Delphi technique eliminates potential bias due to individual group dynamics and is financially feasible.

Validation measures for triage scales

An accepted performance indicator for measuring validity of triage scales is sensitivity, which measures inherent characteristics of the scale with percentage of over-/under-triage against a reference standard. To our knowledge no accepted norms exist for over-/under-triage at each acuity level. The accepted range for average under-triage of not more than 5%-10%, which the ACSCOT considers unavoidable and an associated average over-triage rate of 30%-50% applies exclusively to trauma patients and therefore has limited use when assessing a mixed patient population of trauma and non-trauma cases.⁷⁹ A recent retrospective cohort study indicated that achieving these ACSCOT benchmarks was not feasible in Pennsylvania, and that these guidelines needed modification if they were to be implemented.⁸⁰ This suggests that the accepted ACSCOT ranges are not evidence based and are not suitable for an EC setting where trauma and medical cases are seen together.

Receiver operating curve characteristic curves and multiple logistic regression were considered as alternative statistical validation methods. Neither were deemed appropriate due to the nature of the scale with both continuous and categorical data as well as a list of clinical discriminators and the inherent challenge of not having a gold standard for acuity.

The findings of this thesis show that the average sensitivity over all acuity levels was 75% and specificity 91%. Average under-triage (10%) and over-triage (15%) with respect to the true acuity fell within the accepted ACSCOT ranges, and confirms previous studies¹⁰ that have shown similar average under-triage (7.8%) and over-triage (13.6%) with respect to the true acuity on SATS performance used among South African nurses. The results of this study indicate that the SATS has good performance characteristics and is a valid scale, which may be implemented further in similar settings in South Africa.

Sensitivity is defined as the percentage of vignettes thought to be of a given acuity by raters among all identified by the experts as truly having that acuity. Under- triage for each acuity level is defined as the percentage of vignettes with that “true” acuity that received a less urgent acuity rating from the raters. Percent over-triage of a given acuity level is the percentage of vignettes with that “true” acuity that received a more urgent acuity rating from the raters. Specificity is defined as the percentage of vignettes thought not to be of a given acuity by raters among all identified as truly not having that acuity by the experts. Sensitivity, specificity and percentage over-/under-triage are accuracy summaries (accuracy defined here as the distance from the truth). These summaries may be compared across studies and do not depend on the acuity distribution in a given setting. By contrast, predictive values and over-/under-prediction provide information on actual performance in a given setting. In practice, when a patient is assigned a particular acuity level, the true acuity is not known, and what is of interest from the patient care and resource management perspective is how likely that assigned acuity is to be correct. In this thesis, the acuities in the set of vignettes represented the distribution of EC patients at a busy urban hospital, where the acuity distribution was skewed towards emergency and very urgent acuities.

Consequences of mis-triage versus mis-prediction

Under-triage and under-prediction are a concern to patient care, implying longer waiting times, delayed definitive patient care, leading to increased mortality and morbidity. Over-triage and over-prediction do not directly impinge on patient care, but may indirectly compromise patient care for the collective because overstretched and limited resources are diverted from those in genuine need that are truly a higher priority. Over-triage and over-prediction are therefore important considerations in resource poor settings, where resource allocation, if inappropriately prioritized, may lead to loss of life. Furthermore over-triage/-prediction may result in an over-utilization of finite resources and create an unnecessary financial burden to the health care system, as has been reported in other countries.⁹³ While the extent of under- and over-triage reflects the accuracy of the SATS implementation in a way that is comparable across settings, it is the extent of under- and over-prediction that represents the actual potential impact on patient care and resources in a given setting.

The vignettes in this thesis reflect the distribution of acuities in the population attending secondary level hospitals, where 53% were considered emergency or very urgent by the reference standard, requiring

admission. This is similar to previous studies at other urban secondary level hospitals that also have 50% of their EC presentations being admitted as emergency or very urgent patients.¹⁰

Table 4 presents two hypothetical contexts and this study context with different acuity distributions. **Figure 3** is derived from the three contexts in **Table 4** and illustrates that the percentages of over and under-prediction depend upon the acuity distribution. In a primary care setting, where the distribution of acuity is skewed towards urgent and routine, this triage scale will tend towards greater mis-prediction of emergency and very urgent acuity, and will have higher predictive value for urgent and routine acuity. With higher predictive value at urgent and routine levels of acuity typical of primary care settings, the major impact of poor performance is over-prediction of emergency and very urgent acuity levels, which may result in wastage of scarce and precious resources.

	% acuity distribution		
	A*	B**	C***
emergency	5	40	8
very urgent	5	50	45
urgent	50	5	33
routine	40	5	14

Table 4: Two hypothetical contexts and this study with different acuity distributions⁹⁴

*context A: acuity distribution skewed towards urgent and routine categories

** context B: acuity distribution skewed towards emergency and very urgent categories

*** context C: acuity distribution in this study

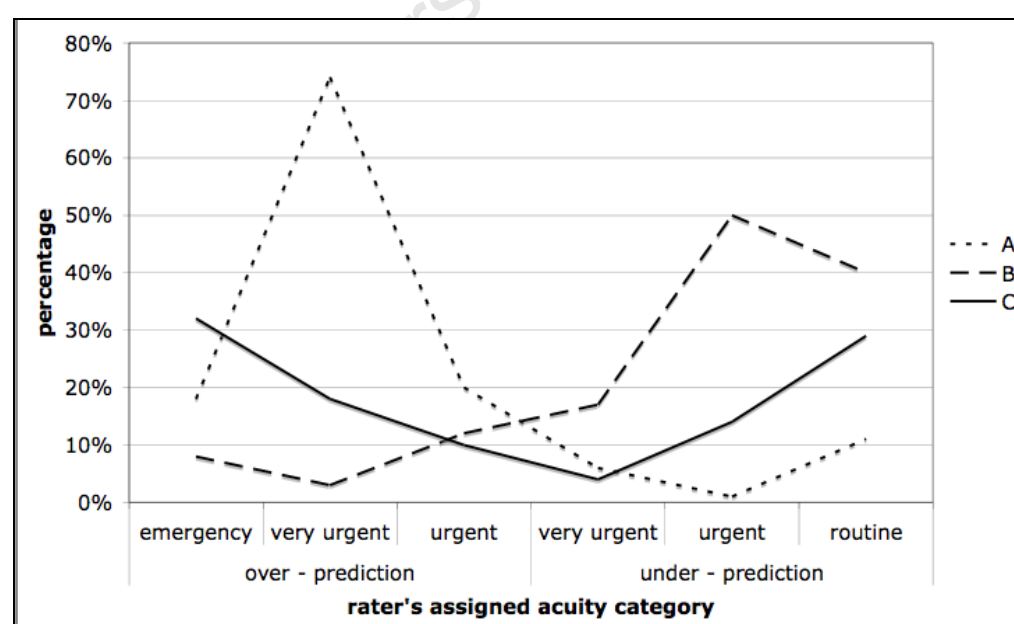


Figure 3: Illustration of over- and under-prediction of rater's assigned acuity categories as acuity distribution changes in different contexts A, B and C⁹⁴

Conversely, as one moves higher up the health care referral chain to secondary, tertiary and highly specialised quaternary levels of care, the triage tool will have higher predictive value at emergency and very urgent acuity levels, and less so for urgent and routine acuities emphasising the desirability of appropriate access at all times to the correct levels of care in an optimally and rationally organised health care system.

The findings of this thesis provide an appropriate alternative and added method for triage scale validation and demonstrate that the construct of acuity as measured with the SATS is valid. By utilizing the set of reference vignettes developed via modified Delphi consensus method, we may be able to perform more comparative studies on triage scales in less developed countries and overcome some of the common barriers (such as poor record keeping and resource limitations) experienced.

Limitations and gaps in the thesis

Reliability

Reliability studies in a real-time environment are challenging and have an ethical limitation in that patient care may be compromised and waiting times for others increased as patients are triaged twice in an already resource limited setting. Hence the findings of this thesis only represent a small pilot in a real-time environment. No further sites were included due to the inherent difficulties of burdening an already resource limited setting.

A limitation in using vignettes is that many nonverbal and contextual cues may be missed despite the fact that written vignettes are a very cost efficient way of assessing reliability.⁹⁵⁻⁹⁷ To overcome the lack of qualitative nonverbal cues the subjective information was retained in the description of vignettes, while keeping the patient information anonymous.

Validation

The reference ranges for average over- and under-triage provided by the ACSCOT are a limited way of interpreting the standard validity indicators as the reference ranges were developed for pre-hospital trauma triage, they only pertain to average over/under-triage and not to each acuity level, and there are no reference ranges for predictive values.

Limitations of the Delphi method are mostly a result of poorly conducted studies rather than fundamental problems with the methodology. Response rates can be low and may decrease in the second round.⁹⁸

However non-response and attrition did not occur in this thesis as each Delphi panellist personally gave assurance of their participation. The selection of the Delphi panel was determined by the research question and the panel may have been limited in its representativeness, as only those experts with an interest may have become involved as participants. Experts from a nursing and emergency physician background were invited from developed and less developed countries. However the majority responded from developed countries. Another potential weakness of the Delphi method is that it may overlook important minority issues because it tries to obtain consensus.⁹⁹ This method of using 80% consensus vignettes may also systematically bias the validity estimate towards better performance. However despite these limitations we feel that the Delphi method is an appropriate alternative to develop a reference with which to test the validity of triage scales in developed and less developed countries.

Conclusion and recommendations for further research

This thesis has presented five papers that describe the performance characteristics of the SATS when used in emergency centres in the Western Cape.

The first contribution of this thesis has been to describe the inter- and intra-rater reliability of SATS ratings for ENAs and emergency physicians triaging independently at secondary emergency centre level. It has shown that adequately trained ENAs are as competent in triage as doctors and demonstrate high inter- and intra-rater reliability. Different reliability measures with their limitations were compared and an alternative graphical approach has been shown to provide a more informative reliability summary. Future publications on triage scale reliability should (i) clearly focus on reliability as distinct from validity (ii) have a clear description of the methods used (iii) include the graphical approach to provide a more informative reliability summary.

The second contribution has been to validate the SATS when used at secondary emergency centre level, describing both the acceptable extent of mis-triage and mis-prediction. Going beyond most studies, the thesis was additionally able to validate the construct of acuity as measured by the SATS, using a modified Delphi method with international experts, providing an alternative method of validation for triage scales. This provides important corroboration for the validity findings.

Further research into accepted norms for extent of over-/under-triage and over-/under-prediction at different acuity levels in the hospital context for both trauma and medical problems is recommended to ensure that appropriate reference ranges and guidelines are available for less developed countries. Comparison studies assessing the practical use of the SATS at primary care level versus central hospital

level setting are also recommended to provide insight into the impact that different acuity distributions have on over- and under-prediction.

Appendix: Additional co-authored papers

Papers referenced in the thesis, but not formally included as part of the thesis

The following papers in which the candidate is an author, have been referenced in either the literature review or synopsis, and were contributed to during the period when the thesis was conceptualized or during the period of registration.

1. Wallis LA, Twomey M. Workload casemix in Cape Town emergency departments. SAMJ 2007;97:1276-80.
2. Twomey M, Wallis LA, Thompson ML, Myers JE. Triage-weighted kappa: a more appropriate triage reliability measure. J Clin Epi 2010;63:806-7.
3. Twomey M, Mullan PC, Torrey SB, Wallis LA, Kestler A. The Princess Marina Accident and Emergency Scale (PATS) provides highly reliable triage acuity ratings. Emerg Med J, in press.

References

1. World Bank Data. South African Population. 2009 [cited 2010/10/15]: Available from: <http://data.worldbank.org/country/south-africa>.
2. Goudge J, Cornell J, McIntyre D, et al. Private sector financing. In: Ntuli A, Suleman F, Barron P, McCoy D: editors. South African Health Review 2001. Durban: Health Systems Trust; 2001.
3. Bradshaw D, Groenewald P, Laubscher R, Nannan N, Nojilana B, Norman R, Pieterse D and Schneider M. Initial Burden of Disease Estimates for South Africa, 2000. Cape Town: South African Medical Research Council, 2003.
4. Larrey DJ. Memoirs of Military Surgery, and Campaigns of the French Armies: Vol 2. Hall RW, translator. Baltimore, MD: Joseph Cushing; 1814:123. Reprinted, Classics of Medicine Library; 1987.
5. Richardson RG. Larrey: Surgeon to Napoleon's Imperial Guard. London, England: Quiller Press Ltd. 2002.
6. Blagg CR. Triage: Napoleon to the present day. J Nephrol. 2004;17:629-632.
7. Iserson KV, Moskop JC. Triage in Medicine, Part I: Concept, History, and Types. Ann Emerg Med. 2007;49:275-281.
8. Hodgetts TJ, MackWay-Jones K. Major Incident Medical Management and Support, 2nd edition: Advanced Life Support Group. BMJ Publishing Group, London, 2002.
9. Mackway-Jones K. Emergency triage, 2nd edition: Manchester Triage Group. Blackwell Publishing Ltd, UK, 2006.
10. Bruijns SR, Wallis LA, Burch VC. A prospective evaluation of the Cape triage score in the emergency department of an urban public hospital in South Africa. Emerg Med J. 2008;25:398-402.
11. Policy for implementation of the triage of patients in Western Cape emergency units. Circular H7 of 2006. Department of Health, Provincial Government, 2006.
12. American College of Emergency Physicians Task Force. Hospital and emergency department overcrowding. Ann Emerg Med 1990;19:336.

13. Canadian Association of Emergency Physicians and National Emergency Nurses Affiliation. Joint Position Statement on emergency department overcrowding [policy]. CJEM 2001;3(2):82-4.
14. Miro O, Antonio MT, Jimenez S, De Dios A, Sanchez M, Borrás A, et al. Decreased health care quality associated with emergency department overcrowding. Eur J Emerg Med 1999;6:105-7.
15. Richardson DB. Increase in patient mortality at 10 days associated with emergency department overcrowding. Med J Aust 2006;184:213-16.
16. Spirivulis PC, Da Silva JA, Jacobs IG, Frazer ARL, Jelinek GA. The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. Med J Aust 2006;184:208-12.
17. Wurz RC, Milne LW, Eitel DR, Travers D, Gilboy N. Reliability and Validity of a new five-level triage instrument. Acad Emerg Med 2000;7:236-42.
18. Fernandes CMB, Tanabe P, Gilboy N, Johnson LA, McNair RS, Rosenau AM, Sawchuk P, Thompson DA, Travers DA, Bonalumi N, Suter RE. Five-Level Triage: A Report from the ACEP/ENA Five-Level Triage Task Force. J Emerg N 2005;31:39-50.
19. FitzGerald GJ. Emergency department triage: a thesis accepted for the degree of Doctor of Medicine. Brisbane: University of Queensland, 1990.
20. Beveridge R. CAEP issues. The Canadian Triage and Acuity Scale: a new and critical element in health care reform. Canadian Association of Emergency Physicians. J Emerg Med 1998;16:507-11.
21. Cooke MW, Jinks S. Does the Manchester triage system detect the critically ill? J Accid Emerg Med 1999;16:179-81.
22. Fan J, Al Darrab A, Eva K, Fernandes CM. Triage Scales in the Emergency Department: A Systematic Review. Annals of Emergency Medicine. 2005; 46: S41.
23. Lewis RJ. Reliability and Validity: Meaning and Measurement. Presented at 1999 Annual Meeting of the Society for Academic Emergency Medicine (SAEM) in Boston, Massachusetts. Available from <http://www.ambpeds.org/ReliabilityandValidity.pdf>

24. Australasian College for Emergency Medicine. Policy on the Australasian Triage Scale. 2000 [cited 2010/11/13]: Available from:
http://www.acem.org.au/media/policies_and_guidelines/P06_Aust_Triage_Scale_-_Nov_2000.pdf.
25. Australasian College for Emergency Medicine. Guidelines on the implementation of the Australasian Triage Scale in emergency departments. 2000 [cited 2010/11/13]: Available from:
http://www.acem.org.au/media/policies_and_guidelines/G24_Implementation_ATS.pdf
26. Canadian Association of Emergency Physicians and National Emergency Nurses Affiliation of Canada. Implementation Guidelines for the Canadian emergency department triage and acuity scale (CTAS). 1998 [cited 2010/11/10]: Available from:
<http://www.caep.ca/template.asp?id=B795164082374289BBD9C1C2BF4B8D32#guidelines>.
27. Bullard MJ, Unger B, Spence J, Grafstein E. Revisions to the Canadian Emergency Department Triage and Acuity Scale (CTAS) adult guidelines. CJEM 2008;10:136-42.
28. Gilboy N, Tanabe P, Travers DA, Rosenau AM, Eitel DR. Emergency Severity Index, Version 4: Implementation Handbook. AHRQ Publication No. 05-0046-2, Rockville, MD: Agency for Healthcare Research and Quality. May 2005.
29. Taboulet P, Moreira V, Haas L, Porcher R, Braganca A, Fontaine JP, Poncet MC. Triage with the French Emergency Nurses Classification in Hospital scale: reliability and validity. Eur J Emerg Med 2009;16:61-67.
30. Murray M, Bullard M, Grafstein E. Revisions of the Canadian Emergency Department Triage and Acuity Scale implementation guidelines. CJEM 2004;6:421-27.
31. van der Wulp I, van Baar ME, Schrijvers AJP. Reliability and validity of the Manchester Triage System in a general emergency department patient population in the Netherlands: results of a simulation study. Emerg Med J 2008;25:431-34.
32. Storm-Versloot MN, Ubbink DT, Chin a Choi V, Luitse JSK. Observer agreement of the Manchester triage system and the emergency severity index: a simulation study. Emerg Med J 2009; 26:556-60.
33. Olofsson P, Gellerstedt M, Carlstroem ED. Manchester Triage in Sweden – Interrater reliability and accuracy. Int Emerg Nurs 2009;17:143-8.
34. Choi YF, Wong TW, Lau CC. Triage rapid initial assessment by doctor (TRIAD) improves waiting time and processing time of the emergency department. Emerg Med J 2006;23:262-65.

35. Cronin JG. The introduction of the Manchester triage scale to an emergency department in the Republic of Ireland. *Acc Emerg Nurs* 2003;11:121-25.
36. Goransson K, Ehrenberg A, Marklund B, Ehnfors M. Accuracy and concordance of nurses in emergency department triage. *Scand J Caring Sci* 2005;19:432-38.
37. Jimenez JG, Murray MJ, Beveridge R, Pons Pons J, Cortes EA, Ferrando Garrigos JB, Borrás Ferre M. Implementation of the Canadian Emergency Department Triage and Acuity Scale (CTAS) in the Principality of Andorra: Can triage parameters serve as emergency department quality indicators? *CJEM* 2003;5:315-22.
38. Elshove-Bolk J, Menci F, van Rijswijk BTF, Simons MP, van Vugt AB. Validation of the Emergency Severity Index (ESI) in self-referred patients in a European emergency department. *Emerg Med J* 2007;24:170-74.
39. Grossmann FF, Nickel CH, Christ M, Schneider K, Spirig R, Bingisser R. Transporting clinical tools to new settings: Cultural adaptation and validation of the Emergency Severity Index in German. *Ann Emerg Med* 2010;in press.
40. Kryanou M, Geogiadis G, Tsiuiki A. Implementing the Emergency Severity Index triage system in the homeland of Hippocrates (conference abstract). *J Emerg Nurs* 2005;31:428.
41. van Gerven R, Delooz H, Sermeus W. Systematic triage in the emergency department using the Australian National Triage Scale: a pilot project. *Europ J Emerg Med* 2001;8:3-7.
42. FitzGerald G, Jelinek GA, Scott D, Gerditz MF. Emergency department triage revisited. *Emerg Med J* 2010;27:86-92.
43. Parenti N, Ferrara L, Reggiani MLB, Sangiorgi D, Lenzi T. Reliability and validity of two four-level emergency triage systems. *Europ J Emerg Med* 2009;16:115-20.
44. Rutschmann OT, Kossovsky M, Geissbuehler A, Perneger TV, Vermeulen B, Simon J, Sarasin FP. Interactive triage simulator revealed important variability in both process and outcome of emergency triage. *J Clin Epidemiol* 2006;59:615-621.
45. Lee CH, Kuan JT, Chiu TF, et al. Coverage and appropriateness of the Taiwan Adult Triage Complaint List. *J Taiwan Emerg Med* 2007;9:65-71.

46. Ng CJ, Hsu KH, Kuan JT, Chiu TF, Chen WK, Lin HJ, Bullard MJ, Chen JC. Comparison between Canadian Triage and Acuity Scale and Taiwan Triage System in Emergency Departments. *J Formos Med Assoc* 2010;109:828-37.
47. Travers DA, Waller AE, Bowling JM, Flowers D, Tintinalli J. Five-level triage system more effective than three level in tertiary emergency department. *J Emerg Nurs* 2002;28:395-400.
48. Streiner DL, Norman GR. *Health Measurement Scales*. 4th ed. Oxford University Press; 2008.
49. Nelson JC, Pepe MS. Statistical description of interrater variability in ordinal ratings. *Statistical Methods in Medical Research* 2000;9:475-96.
50. Jakobsson U, Westergren A. Statistical methods for assessing agreement for ordinal data. *Scand J Caring Sci* 2005;19:427-31.
51. Cohen J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 1960;20:37-46.
52. Fisher RA. *Statistical methods for research workers*. Oliver and Boyd, Edinburgh;1925.
53. Fleiss J. *Statistical methods for rates and proportions*. 2nd ed New York (NY):Wiley;1981.
54. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin* 1968;70:213-20.
55. Cicchetti D, Allison T. A new procedure for assessing reliability of scoring eeg sleep recordings. *American Journal EEG Technology* 1971;11:101-09.
56. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measure of reliability. *Educational and Psychological Measurement* 1973;33:613-19.
57. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 1979; 86:420-428.
58. Brenner H, Kliebsch U. Dependence of weighted kappa coefficients on the number of categories. *Epidemiology* 1996;7:199-202.
59. Ker M. Issues in the use of kappa. *Investigative Radiology* 1991;26:78-83.

60. Maclure M, Willett WC. Misinterpretation and misuse of the kappa statistic. *American J Epidemiology* 1987;126: 161-169.
61. Guggenmos-Holzmam I. How reliable are chance-corrected measures of agreement? *Statistics in Medicine* 1993;12:2191-205.
62. Feinstein AR, Cicchetti DV. High agreement but low kappa. *Journal of Clinical Epidemiology* 1990;43:543-49, 553-58.
63. Landis JR, Koch GG. The measurement of agreement for categorical data. *Biometrics* 1977;33:159-74.
64. Cicchetti DV, Sparrow SA. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behaviour. *American Journal of Mental Deficiency* 1981;86:127-37.
65. Moll HA. Challenges in the validation of triage systems at emergency departments. *J Clin Epi* 2010;63:384-88.
66. Streiner DL, Norman GR. "Precision" and "Accuracy": Two terms that are neither. *J Clin Epi* 2006;59:327-30.
67. Terre Blanche M, Durrheim K. *Research in Practice: Applied Methods for the Social Sciences*. University of Cape Town Press;1999.
68. Messick S. Test validity and the ethics of assessment. *American Psychologist* 1980;35:1012-27.
69. Cronbach LJ, Meehl PE. Construct validity in psychological tests. *Psychological Bulletin* 1955;52:281-302.
70. Clark LA, Watson D. Constructing validity: Basic issues in objective scale development. *Psychological Assessment* 1995;7:309-19.
71. Altman DG, Bland JM. Diagnostic tests 2: predictive values. *BMJ* 1994;309:102.
72. Wuertz RC, Travers D, Gilboy N, Eitel DR, Rosenau A, Yazhari R. Implementation and Refinement of the Emergency Severity Index. *Acad Emerg Med* 2001;8:170-76.

73. Eitel DR, Travers DA, Rosenau AM, Gilboy N, Wuerz RC. The Emergency Severity Index Triage Algorithm Version 2 is Reliable and Valid. *Acad Emerg Med* 2003;10:1070-80.
74. Tanabe P, Gimbel R, Yarnold PR, Adams JG. The Emergency Severity Index (version 3) 5-level Triage System Scores Predict ED Resource Consumption. *J Emerg Nurs* 2004;30:22-29.
75. Tanabe P, Gimbel R, Yarnold PR, Kyriacou DN, Adams JG. Reliability and Validity of Scores on the Emergency Severity Index Version 3. *Acad Emerg Med* 2004;11:59-65.
76. Wollaston A, Fahey P, McKay M, Hegney D, Miller P, Wollaston J. Reliability and validity of the Toowoomba adult trauma triage tool: a Queensland, Australia study. *Acc Emerg Nurs* 2004;12:230-237.
77. Considine J, LeVasseur SA, Villanueva E. The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Ann Emerg Med* 2004;44:516-523.
78. Christian M, Farmer J, Young B. Disaster Triage and Allocation of Scarce Resources. In: Geiling J, editor. *Fundamental disaster medicine*. Mount Prospect (IL): Society of Critical Care Medicine, 2009.
79. American College of Surgeons Committee on Trauma. Resources for optimal care of the injured patient: 1999 Chicago: American College of Surgeons, 1998.
80. Mohan D, Rodengart MR, Farris C, et al. Assessing the feasibility of the American College of Surgeons' Benchmarks for the triage of trauma patients. *Arch Surg* 2011; 43: E1-E7.
81. Wallis LA, Twomey M. Workload casemix in Cape Town emergency departments. *SAMJ*;97:1276-80.
82. Gottschalk S, Wood D, De Vries S, Wallis L, Bruijns S, On behalf of the Cape Triage Group . The cape triage score: a new triage system South Africa. Proposal from the cape triage group. *Emerg Med J* 2006;23:149-153.
83. Bruijns SR, Wallis LA, Burch VC. Effect of introduction of nurse triage on waiting times in a South African emergency department. *Emerg Med J* 2008;25:395-397.
84. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, Roberts C, Shoukri M, Streiner DL. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *International Journal of Nursing Studies* 2011;48:661-671.

85. AGREE for Windows [computer program]. Version 7.002 1999. University of Groningen. Available from URL:
[URL:http://www.scienceplus.nl/component/page,shop.product_details/flypage,shop.flypage_splus/product_id,59/category_id,13/manufacture_id,0/option,com_virtuemart/Itemid,26/](http://www.scienceplus.nl/component/page,shop.product_details/flypage,shop.flypage_splus/product_id,59/category_id,13/manufacture_id,0/option,com_virtuemart/Itemid,26/)
86. van der Wulp I, van Stel F. Adjusted weighted kappa for severity of mistriage decreases reported reliability of emergency department triage systems: a comparative study. *J Clin Epidemiol* 2009;62:1196-1201.
87. Twomey M, Mullan PC, Torrey SB, Wallis LA, Kestler A. The Princess Marina Accident and Emergency Scale (PATS) provides highly reliable triage acuity ratings. *Emerg Med J*. In press.
88. Razzak, Junaid A. and Kellermann, Arthur L. Emergency medical care in developing countries: is it worthwhile? *Bull World Health Organ*, 2002; 80: 900-905.
89. Robertson MA, Molyneux EM. Triage in the developing world – can it be done? *Arch Dis Child*. 2001; 85: 208-213.
90. Personal Communication with Emergency Doctor Patriacia Neto, Quinta D’or Hospital in Rio de Janeiro. May, 2007.
91. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ*. 1995; 311: 376-80.
92. Wallis LA, Carley S, and Hodgetts CT. A procedure based alternative to the injury severity score for major incident triage of children: results of a Delphi consensus process. *Emerg. Med. J*. 2006; 23: 291-95.
93. Carr J. The cost of over-triage on our nation’s health system. 10/09/2010 13:16:00. Available from URL
http://www.healthcanal.com/public-health-safety/10597-The-cost-over-triage-our-nations-health-system.html?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+healthnewshc%2FOxfp+%28Health+News+from+HealthCanal.com%29
94. Personal Communication with Research Professor Mary Lou Thompson. August 2011.
95. Worster A, Sardo A, Eva K, Fernandes CM. Triage Tool Inter-rater Reliability: A Comparison of Live Versus Paper Case Scenarios. *Journal of Emergency Nursing* 2007;33:319-323.

96. Worster A, Fernandes CM. Paper case scenarios in the assessment of triage tools. *Annals of Emergency Medicine*. 2004;44:524-526.
97. Considine J, LeVasseur SA, Villanueva E. The Australasian Triage Scale: examining emergency department nurses' performance using computer and paper scenarios. *Annals of Emergency Medicine*. 2004;44:516-523.
98. Okoli C, Pawlowski S. The Delphi method as a research tool: an example, design considerations and applications. *Information and Management*. 2004;42:15-29.
99. Van Teijlingen E, Pitchforth E, Bishop C, Russell E. Delphi Method and Nominal Group Techniques in Family Planning and Reproductive Health Research. *Journal of Family Planning and Reproductive Health Care*. 2006; 32: 249-252.
100. Twomey M, Wallis LA, Thompson ML, Myers JE. Triage-weighted kappa: a more appropriate triage reliability measure. *J Clin Epi* 2010;63:806-7.
101. Sun JH, Wallis LA, Twomey M, Tran J. The need for a usable methodology to analyze the efficacy of emergency care systems in developing countries. *Acad Emerg Med*. Submitted

